Technische Universität München

Zentrum Mathematik

CARMA-stable models for forward prices

Masterthesis

of

Hannes Hoffmann

Themenstellerin:	Frau Prof. Dr. Klüppelberg
Betreuerin:	Frau Prof. Dr. Klüppelberg
Abgabetermin:	04.01.2013

Hiermit erkläre ich, dass ich die Masterarbeit selbstständig angefertigt und nur die angegebenen Quellen verwendet habe.

München, den 4. Januar 2012

Contents

1 Preface

The major aim of this thesis is to propose a multivariate model for energy futures. Due to the liberalization of energy markets across Europe in the late nineties of the last century, there is an increasing interest in options on energy products. Many of these options have two energy commodities as underlying, whereas they can be modeled either by two single univariate processes or a multivariate process. Therefore we propose such a multivariate model in this thesis, but our focus will be on the estimation of the process and not on the pricing of the derivatives.
In general there exist two types of models to price futures. The first one is to model the spot price and then to derive the future as the risk neutral expectation of the future spot price. For example Benth et al. followed this approach for electricity futures in [3], where they used a stable CARMA(2,1) process to model the short-term variations of the spot price in conjunction with a deterministic seasonal function and a Lévy process for the long term factor of the spot. From this they derived an explicit representation for the futures price. I will also use the multivariate analogoue of a CARMA process to model the short term variations, which was introduce by Marquardt et al. in [24] a few years ago. The second model which is often used in the pricing of fixed income products, is to model the futures directly. The common framework for this is the Heath-Jarrow-Morton (HJM) model. I will use the second framework in this thesis, since a special characteristic of energy markets, such as the natural gas or electricity market, is, that the spot is actually a future with time to maturity of one day. This is due to the fact that producers can neither store their excess electricity in large quantities nor can they adjust their production to spontaneous increases/decreases in demand therefore the producers have to plan their production levels in advance. Furthermore there are also other reasons like the transmission capacities of the network which do not allow for a continuous market. These reasons result in another characteristic of electricity prices. That is why there can occur very high or even negative price peaks in a short time. I will incorporate this behavior in the model by a stable background-driving noise. Since the multivariate model should also capture different energy commodities, we will allow our driving noise to have different indices of stability. For this reason we will introduce the concept of Lévy copulas which describe the dependency of the Lévy measures.
The outline of the thesis is as follows:
In chapter two I will give a short introduction to Lévy processes and copulas. Furthermore I describe a procedure to construct a multivariate Lévy process with different stable margins and how to simulate samples from this distribution. In chapter three I will introduce MCARMA processes and their relationship to continuous-time linear state space models, whereas I present four explicit state space representations for the MCARMA process. In the following subsection I discuss the estimation procedure for the MCARMA parameters and how to recover the driving Lévy process. Chapter four describes the basic model for the energy futures and how it is related to the MCARMA process. The fifth chapter is a simulation study of the performance of the estimators of chapter three. Finally we present a empirical analysis of the German and Italian spot market in chapter six.

2 Lévy Processes and Copulas

2.1 Lévy processes

Definition 2.1 (Lévy Process)**.** A stochastic process $\{\boldsymbol{L}_t\}_{t\in\mathbb{R}^+}$ is a *Lévy process* if

1. $\boldsymbol{L}_0 = 0$ a.s.
2. $\boldsymbol{L}$ has independent increments
3. $\boldsymbol{L}$ has stationary increments
4. $\boldsymbol{L}$ has càdlàg paths.

Furthermore the characteristic function of a Lévy process has a very simple form, i.e.

$$\varphi_t(\theta) = E\left(\exp(i\theta'\boldsymbol{L}_t)\right) = \exp(t\eta(\theta)), \tag{2.1}$$

where

$$\eta(\theta) = -\frac{1}{2}\theta'\Sigma\theta + i\gamma'\theta + \int_{\mathbb{R}^d\setminus\{0\}} e^{i\theta' x} - 1 - i\theta' x\mathbf{1}_{\{\|x\|\le 1\}}\nu(\mathrm{d}x). \tag{2.2}$$

Here $\gamma \in \mathbb{R}^d, \Sigma$ is a positive semidefinite $d\times d$ matrix and ν is the so called Lévy measure which satisfies $\int_{\mathbb{R}^d\setminus\{0\}}(\|x\|\wedge 1)\nu(\mathrm{d}x) < \infty$. The above representation is known as the Lévy-Khintchine-formula and a proof can be found in every book about Lévy processes, for instance in [28]. It is worth mentioning that every Lévy process is uniquely described by the above parameters (γ,Σ,ν) and therefore they are called the generating triplet of the Lévy process. We can interpret γ as somekind of a drift component, Σ determines the continuous part and ν the jump part of the process. For example in the case of $(\gamma,\Sigma,0)$ the corresponding Lévy process is a Brownian motion with drift. Since we are only interested in pure jump Lévy processes in this work we will focus on the generating triplet $(\gamma,0,\nu)$.
In the following we need a Lévy process which is defined on the whole and not only on the positive real line. To obtain such a stochastic process we take two independent copies $\boldsymbol{L}^{(1)}$ and $\boldsymbol{L}^{(2)}$ of our Lévy process and set

$$\boldsymbol{L}_t := \boldsymbol{L}^{(1)}_t\mathbf{1}_{[0,\infty)}(t) - \boldsymbol{L}^{(2)}_{-t-}\mathbf{1}_{(-\infty,0)}(t),$$

where $\boldsymbol{L}_{t-} = \lim_{x\uparrow t}\boldsymbol{L}_x$.
Next we will state a very useful result which we will need later in the derivation of a discretized process.

Lemma 2.2. *Let $G \in \mathbb{R}^{N\times d}$ be a measurable and bounded function and let $\boldsymbol{L}_t$ be a d-dimensional Lévy process with Lévy symbol $\eta(\theta)$. The characteristic function of $\int_s^t G(u)\,\mathrm{d}\,\boldsymbol{L}_u$ for $0\le s<t$ is then given by*

$$E\left[e^{i\theta'\int_s^t G(u)\,\mathrm{d}\,\boldsymbol{L}_u}\right] = e^{\int_s^t \eta(\theta' G(u))\,\mathrm{d}\,u},$$

with the special cases (if they exist)

$$E\left[\int_s^t G(u)\,\mathrm{d}\,\boldsymbol{L}_u\right] = (-i\nabla\eta(0))\int_s^t G(u)\,\mathrm{d}\,u$$

and

$$\mathrm{Cov}\left[\int_s^t G(u)\,\mathrm{d}\boldsymbol{L}_u\right] = \int_s^t G(u)\left(-\mathrm{Hess}\left(\eta(0)\right)\right)G'(u)\,\mathrm{d}u,$$

where $\nabla\eta(0) = \nabla_\theta\eta(\theta)|_{\theta=0}$ is the gradient of η at the d-dimensional zero and Hess denotes the Hessian matrix.

Proof. First assume that G is a step function, i.e. $G(t) = \sum_{i=1}^n G_i\mathbf{1}_{(t_{i-1},t_i]}$ with $s = t_0 < ... < t_n = t$. Since

$$\int_s^t G(u)\,\mathrm{d}\boldsymbol{L}_u = \sum_{i=1}^n G_i(\boldsymbol{L}_{t_i} - \boldsymbol{L}_{t_{i-1}})$$

we get that

$$\begin{aligned} E\left[e^{i\theta'\int_s^t G(u)\,\mathrm{d}\boldsymbol{L}_u}\right] &= \prod_{i=1}^n E\left[e^{i\theta' G_i(\boldsymbol{L}_{t_i}-\boldsymbol{L}_{t_{i-1}})}\right] \\ &= \prod_{i=1}^n e^{(t_i-t_{i-1})\eta(\theta' G_i)} \\ &= e^{\int_s^t \eta(\theta' G(u))\,\mathrm{d}u}, \end{aligned}$$

where we used the stationary increments property of $\boldsymbol{L}$ in the second equation. For the more general result we have to continue the measure theoretic induction, but this is a straightforward calculation. Since

$$(-i)^k \frac{\partial^k\varphi(z)}{\partial z_1^{k_1}\cdots\partial z_N^{k_N}}\bigg|_{z=0} = E(X_1^{k_1}\cdot ...\cdot X_N^{k_n})$$

where $k = \sum_{i=1}^N k_i$, φ is the characteristic function of the random vector $\boldsymbol{X}$ and, because $\eta(0) = 0$, the two other equations are a straightforward differentiation of the first. □

Example 2.3 (Univariate stable process)**.** The Lévy measure of a univariate α-stable process is given by

$$\nu_\alpha(\mathrm{d}x) = \frac{c^+}{x^{1+\alpha}}\mathbf{1}_{\{x>0\}}\,\mathrm{d}x + \frac{c^-}{|x|^{1+\alpha}}\mathbf{1}_{\{x<0\}}\,\mathrm{d}x, \tag{2.3}$$

where $c^+, c^- \geq 0$ and $\alpha \in (0,2)$. Therefore the characteristic function is given as in (2.1) with the generating triplet $(\gamma, 0, \nu_\alpha)$.

Note that if $\alpha = 2$, we still have a stable process, but this corresponds to a Brownian motion, which has continuous paths. The parameter α controls the shape of the tails and c^+, c^- the "skewness" of the distribution, e.g. we observe that ν_α is symmetric iff $c^+ = c^-$ or that our stable Lévy process is a subordinator, i.e. has only increasing sample paths, iff $c^- = 0$. The reason why we put skewness in quotation marks is that an α-stable process has only moments up to order α and hence, excluding the Brownian motion, we do not even have a finite variance, not to mention a third moment.
We could easily extent the univariate stable process to its multivariate counterpart, cf. [26]. In the mentioned paper the Lévy measure of the multivariate stable process is formulated in

terms of its polar coordinates. Therefore it is possible to decompose it into a finite measure σ, which depends exclusively on the direction, and a radial component r, i.e.

$$\nu_\alpha(\mathrm{d}\, r, \mathrm{d}\, u) = \frac{1}{r^{1+\alpha}}\, \mathrm{d}\, r \sigma(\mathrm{d}\, u).$$

The big advantage of this representation is that we can construct an equivalent measure in a comparable way to the univariate case, simply by tempering the Lévy measure as the distance to the origin r increases. Nevertheless we will follow a different approach, since the above decomposition has two major drawbacks. Firstly it is quite hard to estimate the finite measure σ from sample data. Secondly Theorem 2.1.2 [29] states that every linear combination of the components of an α-stable random vector is an α-stable random variable and hence all components must be α-stable. But this assumption seems to be inappropriate in our framework, since we want to model two different commodities or markets. Therefore we desire a random vector whose components have at least the possibility of having distinct indices of stability. Before we obtain such a random vector, we will briefly introduce the notion of Lévy copulas.

2.2 Lévy copulas

The basic idea for Lévy copulas arises from statistical copulas which decompose a multivariate distribution H into a statistical copula C and its corresponding marginal distributions $F^{(i)}$ via

$$H(x_1, ..., x_d) = C(F^{(1)}(x_1), ..., F^{(d)}(x_d)).$$

We could use statistical copulas to get a multivariate distribution which has α_i-stable marginal distributions, but these multivariate distributions are in general not infinitely divisible and hence the resulting process cannot be a Lévy process. Therefore we will use a Lévy copula instead, which ensures that the multivariate process is a Lévy process. The idea is to couple the marginal tail integrals of the Lévy measure and not the marginal distributions.

Definition 2.4 (Lévy copulas)**.** The function $C : \overline{\mathbb{R}}^d \to \overline{\mathbb{R}}$ is a Lévy copula if and only if

1. $C(x_1, ..., x_d) = 0$ if $x_i = 0$ for at least one $i \in \{1, ..., d\}$
2. $C(x_1, ..., x_d) \neq \infty$ for $(x_1, ..., x_d) \neq (\infty, ..., \infty)$
3. C is d-increasing, i.e.

$$V_C((a, b]) = \sum_{x \in \{a_1, b_1\} \times ... \times \{a_d, b_d\}} (-1)^{N(x)} C(x) \geq 0 \ \forall a, b \in \overline{\mathbb{R}}^d,$$

   where $N(x) := \#\{k | x_k = a_k\}$

4. $C^{(i)}(x_i) = x_i \ \forall i \in \{1, ..., d\}, x_i \in \mathbb{R}$, where

$$C^{(i)}(x_i) := \lim_{a \to \infty} \sum_{(x_k)_{k \neq i} \in \{-a, \infty\}^{d-1}} C(x_1, ..., x_d) \prod_{j \neq i} \mathrm{sgn}(x_j).$$

Note that this definition is in line with the definition of a statistical copula, adapted to the fact that the Lévy copula has a domain which is much larger than the unit cube. Therefore it is not surprising that we get an analogous result to the main theorem of statistical copulas. But before we can state it, we have to define an equivalent to the marginal distributions for Lévy measures.

Definition 2.5 (Tail Integrals)**.** Let $\mathcal{I}(x)$ be the tail interval for every $x \in \mathbb{R}$, i.e.

$$\mathcal{I}(x) := \begin{cases} (x, \infty), x \geq 0 \\ (-\infty, x], x < 0. \end{cases}$$

Then the tail integral of a possibly multivariate Lévy measure is given by

$$\overline{\nu}(x_1, ..., x_d) := \prod_{i=1}^{d} \operatorname{sgn}(x_i) \nu \left(\bigtimes_{j=1}^{d} \mathcal{I}(x_j) \right), \text{ for every } x \in \mathbb{R}^d \backslash \{0\}.$$

In the case of a univariate Lévy measure this reduces to

$$\overline{\nu}(x) = \nu\left((x, \infty)\right) \mathbf{1}_{\{x>0\}}(x) - \nu\left((-\infty, x]\right) \mathbf{1}_{\{x<0\}}(x), \text{ for every } x \in \mathbb{R} \backslash \{0\}.$$

Theorem 2.6 (Sklar's theorem for Lévy copulas)**.** *Let $\boldsymbol{L} = (L_1, ..., L_d)$ be a real valued d-dimensional Lévy process. Then there exists a Lévy copula C s.t. the tail integral $\overline{\nu}$ of $\boldsymbol{L}$ and the marginal tail integrals $\overline{\nu}^{(i)}$ of $L_i, i = 1, ..., d$ satisfy*

$$\overline{\nu}(x_1, ..., x_d) = C\left(\overline{\nu}^{(1)}(x_1), ..., \overline{\nu}^{(d)}(x_d)\right), \text{ for all } x \in \mathbb{R}^d \backslash \{0\}. \tag{2.4}$$

The Lévy copula is unique on $\bigtimes_{i=1}^{d} \overline{Ran\ \overline{\nu}^{(i)}}$.
On the other hand if we have a d-dimensional Lévy copula C and d univariate tail integrals $\overline{\nu}^i$ of real valued Lévy processes L_i. Then there exists a d-dimensional Lévy process $\boldsymbol{L}$ whose components are given by L_i and whose tail integral, and hence its Lévy measure, is given by (2.4).

A proof for this theorem can be found in [23]. There the theorem is stated more generally, i.e. they show that (2.4) is also valid for all dimensional reduced tail integrals and their corresponding Lévy copula, but the above theorem will suffice for our purpose. It can easily be seen that if the tail integral in (2.4) is absolutely continuous, then we can recover the corresponding Lévy density via

$$\nu(\mathrm{d}\, x_1, ..., \mathrm{d}\, x_d) = \left. \frac{\partial^d C(u_1, ..., u_d)}{\partial u_1 ... \partial u_d} \right|_{u_1 = \overline{\nu}_1(x_1), ..., u_d = \overline{\nu}_d(x_d)} \nu_1(\mathrm{d}\, x_1) \cdots \nu_d(\mathrm{d}\, x_d).$$

Although we could use every Lévy copula to obtain a multivariate Lévy process from marginal Lévy processes, our next result, which is also adapted from [23], shows that in the case of multivariate stable processes with stable margins the choice of an adequate Lévy copula is restricted to a certain class.

Lemma 2.7. *Let $\boldsymbol{L}$ be a $\mathbb{R}^d$-valued Lévy process and let $\alpha \in (0,2)$. Then $\boldsymbol{L}$ is α-stable if and only if its components $L_1, ..., L_d$ are α-stable and if it has a homogeneous Lévy copula of order 1, that is*

$$C(tx_1, ..., tx_d) = tC(x_1, ..., x_d), \text{ for every } t > 0, x_1, ..., x_s \in \mathbb{R}.$$

This characterization of the dependence structure within a multivariate stable process allows us to construct a generalization of stable processes with the desired property of distinct indices of stability in the marginals.

Definition 2.8. Let $L_1, ..., L_d$ be univariate stable processes with the generating triplet $(\gamma_i, 0, \nu_i)$, where $\gamma_i \in \mathbb{R}, \alpha_i \in (0,2)$ and $\nu_i(\mathrm{d}\, x) := \nu_{\alpha_i}(\mathrm{d}\, x) = \frac{c_i^+}{x^{1+\alpha_i}}\mathbf{1}_{\{x>0\}}\,\mathrm{d}\, x + \frac{c_i^-}{|x|^{1+\alpha_i}}\mathbf{1}_{\{x<0\}}\,\mathrm{d}\, x$, $i = 1, ..., d$. Furthermore let C be a d-dimensional homogeneous Lévy copula of order 1 which is differentiable in all components. Then we call the process $\boldsymbol{L}$ with generating triplet $(\gamma, 0, \nu)$, where $\gamma = (\gamma_1, ..., \gamma_d)'$ and

$$\nu(\mathrm{d}\, x) = \nu(\mathrm{d}\, x_1, ..., \mathrm{d}\, x_d) = \left.\frac{\partial^d C(u_1, ..., u_d)}{\partial u_1 ... \partial u_d}\right|_{u_1 = \overline{\nu}_1(x_1), ..., u_d = \overline{\nu}_d(x_d)} \nu_1(\mathrm{d}\, x_1) \cdots \nu_d(\mathrm{d}\, x_d) \tag{2.5}$$

an $\boldsymbol{\alpha}$-copula-stable process.

Now we want to give an example for a Lévy copula which fulfills the requirements of the last definition.

Example 2.9 (Bidirectional Clayton Copula)**.** For $\delta \in \mathbb{R}^+$ the Clayton Lévy copula is given by

$$C(x, y) = \left[|x|^{-\delta} + |y|^{-\delta}\right]^{-1/\delta}, x, y \in \mathbb{R}_0^+.$$

It can easily be seen that this copula is homogeneous of order one. Furthermore as δ approaches zero, the copula converges to a Lévy copula which corresponds to a multivariate distribution with independent components , whereas for the upper boundary of δ it converges to a Lévy copula of complete dependence. We used the absolute values in the upper definition, because the Clayton copula is also a Lévy copula if its arguments are located in the third quadrant, whereas for the second and fourth it is not. Lévy copulas which have the same domain as the Clayton copula are called positive Lévy copulas. We can construct Lévy copulas which are defined on the whole of $\overline{\mathbb{R}}^2$ by combining positive Lévy copulas and treating every quadrant separately as proposed in [12], that is

$$\begin{aligned} C(x, y) = \quad & C^{++}(a_1|x|, a_2|y|)\mathbf{1}_{\{x\geq 0, y\geq 0\}} - C^{+-}((1-a_1)|x|, (1-a_4)|y|)\mathbf{1}_{\{x\geq 0, y\leq 0\}} \\ & -C^{-+}((1-a_3)|x|, (1-a_2)|y|)\mathbf{1}_{\{x\leq 0, y\geq 0\}} + C^{--}(a_3|x|, a_4|y|)\mathbf{1}_{\{x\leq 0, y\leq 0\}}, \end{aligned}$$

where $a_i \in [0,1]$, $i = 1, ..., 4$ and the C's appearing on the right handside are positive Lévy copulas. Recalling Sklar's theorem, we observe that, for example if $x, y \geq 0$, which corresponds to the case of two upward jumps, then the marginal Lévy measures are $a_1\nu_1(x)\mathbf{1}_{\{x\geq 0\}}$ and $a_2\nu_2(y)\mathbf{1}_{\{y\geq 0\}}$. If we now plug in the Lévy measure of a stable process as in Def. 2.8, we see that $a_1, ..., a_4$ can be chosen arbitrarily since they only affect one of the corresponding scaling

parameters $c_i^{\pm}$. Nevertheless, if $a_1 = ... = a_4 = \eta \in [0,1]$, we get a bidirectional Clayton Lévy copula given by

$$C(x,y) = \left[|x|^{-\delta} + |y|^{-\delta}\right]^{-1/\delta} \left(\eta \mathbf{1}_{\{xy>0\}} - (1-\eta)\mathbf{1}_{\{xy<0\}}\right).$$

To obtain the copula-stable process we have to differentiate this Lévy copula w.r.t. to its two arguments.

$$\frac{\partial^d C(x,y)}{\partial x \partial y} = \operatorname{sgn}(xy)(1+\delta)|xy|^{\delta} \left(|x|^{\delta} + |y|^{\delta}\right)^{\frac{-2\delta-1}{\delta}} \rho(x,y)$$

and hence the joint Lévy measure is

$$\nu(x,y) = \operatorname{sgn}(xy)(1+\delta)|\overline{\nu}_1(x)\overline{\nu}_2(x)|^{\delta} \left(|\overline{\nu}_1(x)|^{\delta} + |\overline{\nu}_2(y)|^{\delta}\right)^{\frac{-2\delta-1}{\delta}} \rho(x,y)\nu_1(x)\nu_2(y),$$

where $\rho(x,y) := \eta \mathbf{1}_{\{xy>0\}} - (1-\eta)\eta \mathbf{1}_{\{xy<0\}}$, ν_i are the marginal Lévy measures of stable processes as in Definition 2.8 with the corresponding tail integrals

$$\overline{\nu}_i(x_i) = \frac{c_i^+}{\alpha_i} x_i^{-\alpha_i} \mathbf{1}_{\{x_i>0\}} - \frac{c_i^-}{\alpha_i} |x_i|^{-\alpha_i} \mathbf{1}_{\{x_i<0\}}.$$

Since we have now a process with the desired properties, the next question is if we can find a measure change which does not distort our process to much - it should remain at least a Lévy process under an equivalent measure. The measure change which naturally arises in this context is the Esscher transform.

Definition 2.10 (Multivariate Esscher transform)**.** Let $\boldsymbol{L}_t$ be a real valued d-dimensional Lévy process on $(\Omega, \mathcal{F}, \{\mathcal{F}\}_t, \mathcal{P})$ with the generating triplet $(\gamma, 0, \nu)$, then we can construct a local equivalent measure $\mathcal{Q}$ for our physical measure $\mathcal{P}$ via the Radon-Nikodym derivative

$$\frac{\mathrm{d}\,\mathcal{Q}|_{\mathcal{F}_t}}{\mathrm{d}\,\mathcal{P}|_{\mathcal{F}_t}} = \exp\left(\theta' \boldsymbol{L}_t - t \ln \eta(-i\theta)\right), \tag{2.6}$$

where $\theta \in \mathbb{R}^d$ and $\eta(\theta) = E\left(\exp(\theta' \boldsymbol{L}_1)\right)$, which is given explicitly in (2.2).

Theorem 2.11. *In the same framework as in the prior definition the process $\boldsymbol{L}_t$ is a Lévy process under $\mathcal{Q}$ with*

$$\gamma_{\mathcal{Q}} = \gamma_{\mathcal{P}} + \int_{0<\|x\|<1} \left(1 - e^{\theta' x}\right) \nu_{\mathcal{P}}(\mathrm{d}\,x)$$

and

$$\nu_{\mathcal{Q}}(\mathrm{d}\,x) = e^{\theta' x} \nu_{\mathcal{P}}(\mathrm{d}\,x).$$

Proof. Let $t \in \mathbb{R}$ then the characteristic function under $\mathcal{Q}$ is given by

$$\begin{aligned}
E_{\mathcal{Q}|\mathcal{F}_t}\left(e^{i\tau' \boldsymbol{L}_t}\right) &= E\left(e^{(i\tau+\theta)'\boldsymbol{L}_t - t\ln\eta(-i\theta)}\right) = E\left(e^{(i\tau+\theta)'\boldsymbol{L}_t}\right) E\left(e^{-t\ln\eta(-i\theta)}\right) \\
&= e^{t(\eta(\tau-i\theta)-\eta(-i\theta))} \\
&= \exp\left[t\left(\gamma_{\mathcal{P}}'(i\tau+\theta) - \int_{\mathbb{R}^d\setminus\{0\}} 1 - e^{(i\tau+\theta)'x} + (i\tau+\theta)'x\mathbf{1}_{\{\|x\|\le 1\}}\nu_{\mathcal{P}}(\mathrm{d}\,x)\right.\right. \\
&\quad \left.\left. -\gamma_{\mathcal{P}}'\theta + \int_{\mathbb{R}^d\setminus\{0\}} 1 - e^{\theta' x} + \theta' x\mathbf{1}_{\{\|x\|\le 1\}}\nu_{\mathcal{P}}(\mathrm{d}\,x)\right)\right] \\
&= \exp\left[t\left(i\gamma_{\mathcal{P}}'\tau + i\int_{0<\|x\|<1} \tau' x\left(1 - e^{\theta' x}\right)\nu_{\mathcal{P}}(\mathrm{d}\,x)\right.\right. \\
&\quad \left.\left. - \int_{\mathbb{R}^d\setminus\{0\}} \left(1 - e^{i\tau' x} + i\tau' x\mathbf{1}_{\{\|x\|\le 1\}}\right)e^{\theta' x}\nu_{\mathcal{P}}(\mathrm{d}\,x)\right)\right].
\end{aligned}$$

By comparing this with (2.1) we obtain the result. □

If we now apply this theorem to the $\boldsymbol{\alpha}$-copula-stable process we get

$$\nu_{\mathcal{Q}}(\mathrm{d}\,x_1,...,\mathrm{d}\,x_d) = \left.\frac{\partial^d C(u_1,...,u_d)}{\partial u_1 ... \partial u_d}\right|_{u_1=\overline{\nu}_1(x_1),...,u_d=\overline{\nu}_d(x_d)} e^{\theta_1 x_1}\nu_1(\mathrm{d}\,x_1)\cdots e^{\theta_d x_d}\nu_d(\mathrm{d}\,x_d),$$

which I will call a $(\boldsymbol{\alpha},\boldsymbol{\theta})$-tempered-copula-stable process.

2.3 Simulation of a copula-stable process

Since we want to validate the estimation procedures for our model, we will need trajectories from the copula-stable process. In general there exist multiple procedures to generate sample paths from a multivariate stable process, but only the one proposed by [32] is suited for our process, since it incorporates Lévy copulas.

Theorem 2.12. *Let ν be a Lévy measure on $\mathbb{R}^d$ with marginal tail integrals of a stable process and a homogeneous Lévy copula $C(x_1,...,x_2)$ of order 1 which is continuous in $(\infty)^d$ and let $\{V_i\}$ be a sequence of independent standard uniform random variables. Furthermore let K be the conditional probability distribution associated to C with distribution function*

$$C_\xi(x_2,...,x_d) = \mathrm{sgn}(\xi)\frac{\partial}{\partial\xi}V_C\left((\min(\xi,0),\max(\xi,0)]\times(-\infty,x_2]\times\cdots\times(-\infty,x_d]\right),$$

where V_C is the volume function as in the definition of a Lévy copula 2.4. Introduce the independent d-dimensional random jump time sequence $\{\Gamma_i\}$, independent of $\{V_i\}$ s.t. $N = \sum_{i=1}^{\infty}\boldsymbol{1}_{\{\Gamma_i^{(1)}\}}$ is a Poisson random measure on $\mathbb{R}$ with intensity measure λ and $\left(\Gamma_i^{(2)},...,\Gamma_i^{(d)}\right)$ conditional on $\Gamma_i^{(1)}$ is distributed on $\mathbb{R}^{d-1}$ with law $K(\Gamma_i^{(1)},\mathrm{d}\,x_2,...,\mathrm{d}\,x_d)$. Finally let

$$A_k(\tau) = \int_{|x|\le 1} x_k\nu_\tau(\mathrm{d}\,x_1\cdots\mathrm{d}\,x_d),\ k = 1,...,d,$$

where

$$\nu_\tau := \mathbf{1}_{\left(-\infty,\overline{\nu}_1^{(-1)}(\tau)\right]\cup\left[\overline{\nu}_1^{(-1)}(\tau),\infty\right)}(x_1)\nu(\mathrm{d}\,x_1,...,\mathrm{d}\,x_d).$$

Then the process $\{\boldsymbol{L}_{\tau,t}\}_{0\le t\le 1}$*, where the k-th component is given by*

$$\boldsymbol{L}_{\tau,t}^{(k)} = \sum_{-\tau\le\Gamma_i^{(1)}\le\tau} \nu_k^{(-1)}\left(\Gamma_i^{(k)}\right)\mathbf{1}_{V_i\le t} - tA_k(\tau),\ k=1,...,d,$$

converges in distribution as $\tau\to\infty$ *to a Lévy process* $\{\boldsymbol{L}_t\}_{t\in[0,1]}$ *with characteristic function*

$$e^{iu'\boldsymbol{L}_t} = \exp\left(t\int_{\mathbb{R}^d}\left(e^{iu'z}-1-iu'z\right)\mathbf{1}_{\{|z|\le 1\}}\nu(\mathrm{d}\,z)\right).$$

Note that it is not necessary to include the drift term $A_k(\tau)$ if the process is of finite variation, but since we assume that the indices of stability $\alpha_i > 1$ for power commodity prices, we have to deal with the infinite variation case. The full proof of the theorem can be found in [32]. Nevertheless the key idea is to truncate the jumps of the process such that they are larger than a common threshold, here $\overline{\nu}_i^{(-1)}(\tau)$. Then we can approximate the original process by a compound Poisson process for which it is easy to generate sample paths. The dependency between the components of the process is modeled by the dependency of the jump times which can be calculated explicitly based on the fact that, given the first jump time, the jump times of the other components are distributed with law K. Finally, since we have truncated all jumps lower than the threshold, we introduce the drift $A_k(\tau)$ which can be viewed as the expected value of the small jumps. It is also possible to refine the approximation of the truncated small jumps by introducing an additional Brownian Motion which mimics the infinite variation of the small jumps. For further details on this approach see [11]. I do not incorporate this refinement, since I am convinced that it would make sense if we were interested in the whole sample path, but our desire is just to model the right end point of the trajectory and therefore I think that the changes under the Brownian Motion are immaterial. With the above theorem we are able to state an algorithm to simulate trajectories of the copula-stable process.

Algorithm 2.13 (Simulation of a two-dimensional copula-stable process)**.** First select a truncation level ε and calculate the drift term $A_i(\overline{\nu}_1(\varepsilon))$, $i=1,2$ as in Theorem 2.12. Evaluate the algorithm below for positive and negative epsilon.

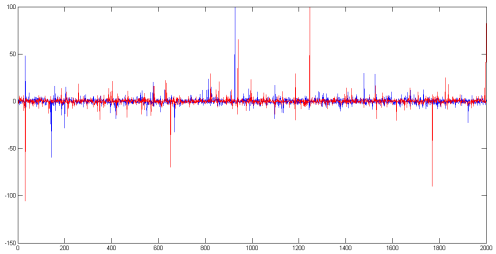
- Set $j=0$, $\Gamma_0^{(1)}=0$, $\tau=\overline{\nu}_1(\varepsilon)$
- IF $\tau=0$
  - GO TO END
- WHILE $|\Gamma_j^{(1)}|<|\tau|$
  - Set $j=j+1$
  - Generate T_j: standard exponential

- $\Gamma_j^{(1)} = \Gamma_{j-1}^{(1)} + \operatorname{sgn}(\varepsilon) T_j$
- Generate U_j and V_j: standard uniform
- $\Gamma_j^{(2)} = C_{\Gamma_j^{(1)}}(U_j)$, where C_ξ is the conditional pdf as in Theorem 2.12

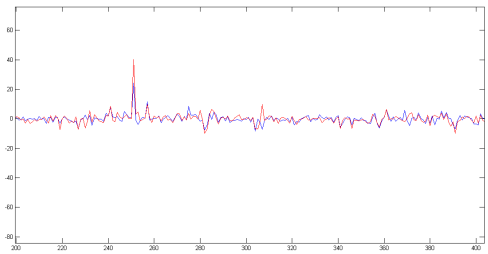
- $L_t^{(i)} = \sum_{k=1}^{j} \mathbf{1}_{V_k \le t} \overline{\nu}_i^{(-1)}(\Gamma_k^{(i)}) - t A_i(|\tau|)$

Finally add up the processes for positive and negative epsilon.

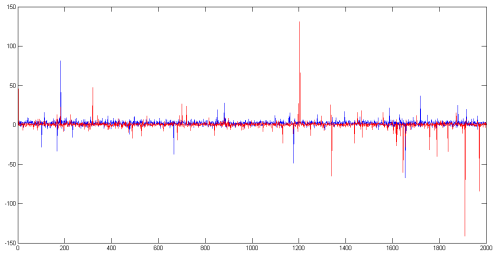
In the above algorithm it is also possible to choose a different jump time sequence $\Gamma^{(1)}$. Here we can name Tankov who proposes in [32] to use the sequence $\Gamma_j^{(1)} = X_j(-1)^j$, where X_j are jump times of a Poisson process with jump intensity 2. I have decided to model the positive and negative jumps separately, since Tankov's sequence has the major drawback that for two consecutive jump times one is always positive and one negative, whereas the positive one is always larger in absolute terms. If we now assume that the first marginal is a unskewed stable process, i.e. $c_1^+ = c_1^-$, then this would result in larger downward jumps compared to the upward jumps, which contradicts the assumption of an unskewed marginal.
Note that there are certain errors inherent to the above stated simulation method. The first is that, due to the truncation, we are not able to model the small jumps appropriately. However this error can be reduced by decreasing the value of ε, such that we get a good balance between the computational burden and the accuracy. The second error arises from the fact that the jump sizes of the second component are not sorted in a decreasing order. Therefore it is also possible that we truncate large jumps which correspond to a jump time of the first component which is above the truncation level. For this reason we face a problem which is far more serious for we cannot control it. This error is most pronounced for Lévy copulas which are close to a copula of independence, e.g. as $\delta \to 0$ in Example 2.9.
In Figure 2.1 I have simulated four different copula-stable increments. The first two graphs show the increments of a copula-stable processes where I have fixed the scale parameters of the Lévy measures of the marginal processes to one, such that the difference between upward and downward jumps is controlled solemnly by η. For the third and fourth graph η is fixed and c^+ and c^- can vary. For each pair one graph corresponds to strong dependence ($\delta = 3 >> 0$) and the other to weak independence ($\delta = 0.3 \approx 0$). We can clearly see that for the strong dependence case the jumps occur at the same time, especially the large ones, whereas at the same time they do not have to jump in the same direction, since $\eta \notin \{0, 1\}$. Moreover I have estimated the marginal parameters via the program Stable by J.P.Nolan (http://academic2.american.edu/ jpnolan/stable/stable.html). The result was that for the first component the algorithm is quite exact, whereas "quite exact" means that the index of stability α was always a bit smaller than the original one, which is due to the fact that we have truncated the stable margin. I also observed the same effect for the second component, but here the approximation of the process worsen as $\delta \to 0$, which was described earlier.



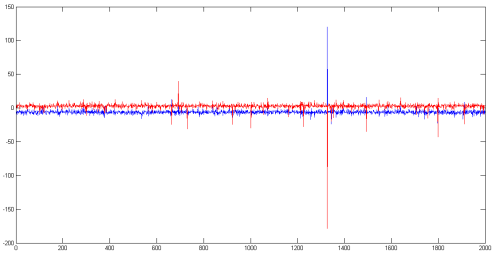
(a) 2000 simulated increments of a $(1.7, 1.5)$-copula stable process with $c^+ = (1,1)$, $c^- = (1,1)$, $\delta = 0.3$ and $\eta = 0.5$, where the second process corresponds to the red curve.



(b) Extract of 2000 simulated increments of a $(1.7, 1.5)$-copula stable process with $c^+ = (1,1)$, $c^- = (1,1)$, $\delta = 3$ and $\eta = 0.75$, where the second process corresponds to the red curve.



(c) 2000 simulated increments of a $(1.7, 1.5)$-copula stable process with $c^+ = (1, 0.75)$, $c^- = (0.75, 1)$, $\delta = 0.3$ and $\eta = 0.5$, where the second process corresponds to the red curve.



(d) Extract of 2000 simulated increments of a $(1.7, 1.5)$-copula stable process with $c^+ = (1, 0.75)$, $c^- = (0.75, 1)$, $\delta = 3$ and $\eta = 0.5$, where the second process corresponds to the red curve.

Figure 2.1: Simulated copula-stable process increments

3 Multivariate CARMA processes and their State Space Representation

3.1 Basic definitions and state space representations

For the rest of this thesis, $\mathbb{R}^{l\times m}[s]$ ($\mathbb{R}^{l\times m}\{s\}$) will denote the set of $l\times m$ dimensional matrices whose entries are polynomial (or rational) functions with real valued coefficients.
CARMA processes are the continuous-time analogue of the well know ARMA processes. For further information on ARMA processes we refer to the comprehensive book of Brockwell and Davis [7]. Although CARMA process have already appeared in the mid forties of the last century in the context of Gaussian processes, they came back into focus in the the many papers of Brockwell, see eg. [8], with a Lévy process as driving process. The extension to CARMA processes of higher dimensions was later investigated by Marquardt and Stelzer [24] a few years ago.
The formal definition of a CARMA process, as well as its multivariate counterpart the MCARMA process, is given in the following definition.

Definition 3.1 (MCARMA Process)**.** Let $\{\boldsymbol{L}_t\}_{t\in\mathbb{R}_+}$ be a d-dimensional Lévy process and let for $0 \le q < p$, $A_1, ..., A_p \in \mathbb{R}^{n\times n}$ and $B_0, ..., B_{q-1} \in \mathbb{R}^{n\times d}$ be the matrix coefficients of the polynomials

$$P(z) := \boldsymbol{I}_n z^p + A_1 z^{p-1} + ... + A_p$$
$$Q(z) := B_0 z^q + B_1 z^{q-1} + ... + B_q.$$

Then the MCARMA(p,q) process $\{\boldsymbol{Y}_t\}$ is given as the solution of the stochastic differential equation

$$P(D)\boldsymbol{Y}_t = Q(D)D\boldsymbol{L}_t,\ t \ge 0, \tag{3.1}$$

where D denotes differentiation with respect to t.

This definition lacks one thing. Take for instance the Brownian motion as the driving Lévy process, then it is know from every introductory course on stochastic processes that the paths of the Brownian motion are nowhere differentiable and hence the derivative $D\boldsymbol{L}_t$ does not exist in the usual sense. Therefore we have to work with a continuous-time linear state space model (clss) which mimics the above stochastic differential equation (SDE). For this purpose we will show that we can find a clss for every MCARMA process s.t. their transfer functions, i.e. the function which connects the driving Lévy process $\{\boldsymbol{L}_t\}_{t\in\mathbb{R}_+}$ and the output process $\{\boldsymbol{Y}_t\}_{t\in\mathbb{R}_+}$, are equal. Hence the transfer function of a MCARMA process is given by $H(z) = P(z)^{-1}Q(z)$.

Definition 3.2 (Continuous-time linear state space models)**.** Let $\boldsymbol{A} \in \mathbb{R}^{N\times N}$, $\boldsymbol{B} \in \mathbb{R}^{N\times d}$, $\boldsymbol{C} \in \mathbb{R}^{n\times N}$ and let $\boldsymbol{L}$ be a d-dimensional Lévy-process. Then the quadruple $(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}, \boldsymbol{L})$ is called an continuous-time linear state space model of dimension N of the stochastic process $\boldsymbol{Y}_t$, if there is a state vector process $\boldsymbol{X} \in \mathbb{R}^N$, which is a Ornstein-Uhlenbeck process given by the state equation

$$\mathrm{d}\,\boldsymbol{X}_t = \boldsymbol{A}\boldsymbol{X}_t\,\mathrm{d}\,t + \boldsymbol{B}\,\mathrm{d}\,\boldsymbol{L}_t$$

and which is connected to $\boldsymbol{Y}$ via the observation equation

$$\boldsymbol{Y}_t = \boldsymbol{C}\boldsymbol{X}_t.$$

The state space model $(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}, \boldsymbol{L})$ of dimension N is called minimal, if there is no other state space model for the process $\boldsymbol{Y}$ with a lower dimension than N. In this case N is called the McMillan degree.

As we will see the state space model for a MCARMA is in general not unique. Therefore the next lemma, which is adapted from [31], will give us the connection between an MCARMA process and a linear state space model.

Lemma 3.3. *If*

$$P^{-1}(z)Q(z) = \boldsymbol{C}(z\boldsymbol{I}_N - \boldsymbol{A})^{-1}\boldsymbol{B}, \tag{3.2}$$

where $(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}, \boldsymbol{L})$ *are given as in the prior definition and* P, Q *are the autoregressive and moving average matrix polynomials, resp., then the MCARMA process, as defined in Definition* 3.1*, and the continuous-time linear state space model yield the same output process* $\boldsymbol{Y}$*.*

Therefore we have to ensure that the transfer function of the clss and of the MCARMA process are equal. As a natural starting point we will set the matrix $\boldsymbol{A}$ to be the first companion matrix of the matrix polynomial $P(z)$, i.e.

$$\boldsymbol{A} = \begin{pmatrix} 0 & \boldsymbol{I}_n & 0 & \dots & 0 \\ 0 & 0 & \boldsymbol{I}_n & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \boldsymbol{I}_n \\ -A_p & -A_{p-1} & -A_{p-2} & \dots & -A_1 \end{pmatrix}.$$

The terminology companion matrix comes from the fact that in the univariate case the characteristic polynomial of $\boldsymbol{A}$, i.e.

$$p_{\boldsymbol{A}}(z) = \det(\boldsymbol{I}_N z - \boldsymbol{A})$$

is equal to the polynom $P(z)$ and hence the eigenvalues of $\boldsymbol{A}$ are the roots of $P(z)$. Unfortunately this relationship only remains the same to a certain extent, that is the eigenvalues of $\boldsymbol{A}$ are the roots of the $\det(P(z))$.
Nevertheless we can calculate $(z\boldsymbol{I}_n - \boldsymbol{A})^{-1}$ manually via a simple Gaussian elimination, which yields

$$(z\boldsymbol{I}_N - \boldsymbol{A})^{-1} =$$

$$\begin{pmatrix} g(p,z)^{-1}g(p-1,z) & g(p,z)^{-1}g(p-2,z) & \cdots & g(p,z)^{-1} \\ -\boldsymbol{I}_n + zg(p,z)^{-1}g(p-1,z) & zg(p,z)^{-1}g(p-2,z) & \cdots & zg(p,z)^{-1} \\ -z\boldsymbol{I}_n + z^2 g(p,z)^{-1}g(p-1,z) & -\boldsymbol{I}_n + z^2 g(p,z)^{-1}g(p-2,z) & \cdots & z^2 g(p,z)^{-1} \\ \vdots & \vdots & \ddots & \vdots \\ -z^{p-2}\boldsymbol{I}_n + z^{p-1}g(p,z)^{-1}g(p-1,z) & -z^{p-3}\boldsymbol{I}_n + z^{p-1}g(p,z)^{-1}g(p-2,z) & \cdots & z^{p-1}g(p,z)^{-1} \end{pmatrix}, \tag{3.3}$$

where

$$g(p,z) := \sum_{i=1}^{p} z^{p-i} A_i + z^p \boldsymbol{I}_n.$$

We immediately observe that $g(p,z)^{-1} = P(z)^{-1}$ and therefore that the last column of (3.3) consists exclusively of our matrix polynomial P and additionally of powers of z up to the order of $p-1$. Therefore we will focus on this last column.
Before we start to formulate various state space representations, we will introduce the Smith-McMillan form for rational matrix functions, cf. [4] Theorem 4.7.5. This canonical decomposition is very important for the first two representation, since on the one hand it diagonalizes a rational matrix function and on the other hand it allows to construct minimal realizations.

Theorem 3.4. *Let $H(z)$ be a $n \times N$ dimensional matrix whose entries are rational functions, i.e $H(z) \in \mathbb{R}^{n\times N}\{z\}$. Furthermore let $r := \operatorname{rank} H(z)$. Then there exist matrices $U_1(z) \in \mathbb{R}^{n\times n}[z]$ and $U_2(z) \in \mathbb{R}^{N\times N}[z]$ with $\det U_i(z) = c_i \in \mathbb{R}\backslash\{0\}$, $i = 1,2$ and polynomials $\varepsilon_1(z),...,\varepsilon_r(z),\psi_1(z),...\psi_r(z) \in \mathbb{R}[z]$ with leading coefficient one s.t. $\varepsilon_i(z)$ and $\psi_i(z)$ are coprime for all $i = 1,...,r$ and $\psi_{i+1}(z)$ divides $\psi_i(z)$ as well as $\varepsilon_i(z)$ divides $\varepsilon_{i+1}(z)$ for all $i = 1,...,r-1$. And we can represent $H(z)$ by*

$$H(z) = U_1(z)M(z)U_2(z), \tag{3.4}$$

where

$$M = \begin{pmatrix} \operatorname{diag}\left(\frac{\varepsilon_i(z)}{\psi_i(z)}\right) & \mathbf{0}_{r\times(N-r)} \\ \mathbf{0}_{(n-r)\times r} & \mathbf{0}_{(n-r)\times(N-r)} \end{pmatrix}. \tag{3.5}$$

The degrees of $\psi_i(z)$ are called Kronecker indices and if $H(z)$ is the transfer function of a clss, then $\sum_{i=1}^{r} \deg \psi_i(z)$ is equal to the McMillan degree.

If we use the multivariate extensions of the "standard" matrices for CARMA processes, i.e.

$$\boldsymbol{C} = (B_q,...,B_0,\mathbf{0},...,\mathbf{0}) \text{ and } \boldsymbol{B} = (\mathbf{0},...,\mathbf{0},\boldsymbol{I}_d)',$$

in conjunction with (3.3), we obtain

$$\boldsymbol{C}(z\boldsymbol{I}_N - \boldsymbol{A})^{-1}\boldsymbol{B} = \sum_{i=0}^{q} z^{q-i} P(z)^{-1} = Q(z)P(z)^{-1}.$$

Unfortunately matrices do not commute in general and thus we do not get exactly the transfer function of the MCARMA process (3.1). Nevertheless we obtain our first state space representation of a MCARMA process.

Lemma 3.5 (Controller canonical representation)**.** *Let $\{\boldsymbol{Y}_t\}$ be the MCARMA process as in Definition 3.1 with matrix polynomials $P(z)$ and $Q(z)$. Then there exist matrix polynomials $P^*(z), Q^*(z)$ and integers $p^* > q^*$ with*

$$P^*(z) := \boldsymbol{I}_n z^{p^*} + A_1^* z^{p^*-1} + ... + A_{p^*}^*$$

$$Q^*(z) := B_0^* z^{q^*} + B_1^* z^{q^*-1} + ... + B_{q^*}^*,$$

s.t.

$$P(z)^{-1}Q(z) = Q^*(z)P^*(z)^{-1}.$$

Furthermore $\det P^*(z) = c \det P(z), c \in \mathbb{R}$ *and* $\{\boldsymbol{Y}_t\}$ *has a state space representation with*

$$\boldsymbol{A} = \begin{pmatrix} 0 & \boldsymbol{I}_d & 0 & ... & 0 \\ 0 & 0 & \boldsymbol{I}_d & ... & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & ... & \boldsymbol{I}_d \\ -A_p^* & -A_{p-1}^* & -A_{p-2}^* & ... & -A_1^* \end{pmatrix}, \boldsymbol{B} = \begin{pmatrix} \boldsymbol{0} \\ \vdots \\ \boldsymbol{0} \\ \boldsymbol{I}_d \end{pmatrix}$$

and

$$\boldsymbol{C} = (B_q, ..., B_0, \boldsymbol{0}, ..., \boldsymbol{0}).$$

Proof. Due to the prior considerations, it remains to prove that such a right matrix fraction description $Q^*(z)P^*(z)^{-1}$ of $P(z)^{-1}Q(z)$ exists. But first we have to ensure that $P(z)$ and $Q(z)$ are left coprime, that is if there exist a matrix polynomial $L(z) \in \mathbb{R}^{n\times n}[z]$ with

$$P(z) = L(z)\widetilde{P}(z) \text{ and } Q(z) = L(z)\widetilde{Q}(z), \tag{3.6}$$

where $\widetilde{P}(z) \in \mathbb{R}^{n\times n}[z]$ and $\widetilde{Q}(z) \in \mathbb{R}^{n\times d}[z]$ are also matrix polynomials, then $\det L(z) = c \in \mathbb{R}\backslash\{0\}$. Matrix polynomials such as $L(z)$ with a constant determinant are called unimodular. Since their determinant is non-zero, they can be inverted via $L(z)^{-1} = \det(L(z))^{-1}L(z)^A$, where $L(z)^A$ is the adjungate matrix of $L(z)$ which is obviously also a polynomial matrix. Therefore we assume that a non unimodular $L(z)$ with highest possible degree exists which fulfills (3.6). Then we know from [27] Theorem 16.14 that

$$(P(z), Q(z))\, U(z) = (L(z), \boldsymbol{0}_{n\times d}),$$

where $U(z) \in \mathbb{R}^{(n+d)\times(n+d)}[z]$ is a unimodular matrix polynomial. Thus by multiplying $U(z)^{-1}$ from the right we get that

$$\det P(z) = \det L(z) \det U_{n,n}(z)^{-1}.$$

Here $U_{n,n}(z)^{-1}$ is the first $n \times n$ submatrix of $U(z)^{-1}$. Therefore the non singularity of $P(z)$ translates to $L(z)$. Due to the fact that $L(z)$ is invertible, we could take the left coprime matrices $\widetilde{Q}(z)$ and $\widetilde{P}(z)$ instead of $P(z)$ and $Q(z)$, since they yield the same transfer function $H(z) := P(z)^{-1}Q(z)$ and as a result of this we will assume in the following that $P(z)$ and $Q(z)$ are left coprime.
Since $H(z) \in \mathbb{R}^{n\times d}\{z\}$, we can use Theorem 3.4, i.e. $H(z) = U_1(z)M(z)U_2(z)$, where $M(z)$ is given as in (3.5). Furthermore we can decompose $M(z)$ a bit further, i.e.

$$M(z) = \Psi_L^{-1}(z)E(z) = E(z)\Psi_R^{-1}(z),$$

where

$$E(z) = \begin{pmatrix} \operatorname{diag}(\varepsilon_i(z)) & \boldsymbol{0}_{r\times(d-r)} \\ \boldsymbol{0}_{(n-r)\times r} & \boldsymbol{0}_{(n-r)\times(d-r)} \end{pmatrix},$$

$$\Psi_L(z) = \begin{pmatrix} \operatorname{diag}(\psi_i(z)) & \mathbf{0}_{r\times(n-r)} \\ \mathbf{0}_{(n-r)\times r} & \boldsymbol{I}_{(n-r)\times(n-r)} \end{pmatrix} \text{ and } \Psi_R(z) = \begin{pmatrix} \operatorname{diag}(\psi_i(z)) & \mathbf{0}_{r\times(d-r)} \\ \mathbf{0}_{(d-r)\times r} & \boldsymbol{I}_{(d-r)\times(d-r)} \end{pmatrix}.$$

Hence we can decompose $H(z)$ into

$$H(z) = (\Psi_L(z)U_1(z)^{-1})^{-1}(E(z)U_2(z))$$

as well as

$$H(z) = (U_1(z)E(z))(U_2(z)^{-1}\Psi_R(z))^{-1}.$$

Finally, due to Proposition 4.7.15 of [4], two left coprime matrix fraction descriptions only differ in a unimodular matrix and hence the equations above are two matrix fraction descriptions of the transfer function $H(z)$ with the desired properties. □

If we want to apply this representation, e.g. for the estimation of the MCARMA parameters, we could formulate our state space representation with the help of the above given right matrix fraction description of the transfer function and convert it afterwards to a left matrix fraction to recover the AR and MA matrix polynomials or the other way around. In either ways we have to convert the matrix fraction description. In general we could proceed as in the last proof, but from a practical point of view this way is a bit cumbersome. Nevertheless there are efficient algorithms in the literature (see eg. [1]). The advantage of this representation is that it splits the MA and AR matrix polynomial coefficients into the state and the observation equation. Furthermore, as we will see later, it allows under certain circumstances to invert the process, that is that we are able to express the state process in terms of the output process. After inversion we are able to recover the driving Lévy noise of the MCARMA process, which we can also use in the estimation of the MCARMA parameters. Due to this property we will see in the data analysis section that this representation performs best for our data set.
Our next state space representation is based on the second companion matrix which is, loosely spoken, the transpose of the first companion matrix.

Lemma 3.6. *Let*

$$\boldsymbol{A} = \begin{pmatrix} 0 & 0 & 0 & -A_p \\ \boldsymbol{I}_n & 0 & 0 & -A_{p-1} \\ \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \boldsymbol{I}_n & -A_1 \end{pmatrix}, \boldsymbol{B} = \begin{pmatrix} B_q \\ \vdots \\ B_0 \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{pmatrix} \text{ and } \boldsymbol{C} = \begin{pmatrix} \mathbf{0} \\ \vdots \\ \mathbf{0} \\ \boldsymbol{I}_n \end{pmatrix}'.$$

Then the continuous-time linear state space representation $(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C})$ is a realization of the MCARMA process as defined in Definition 3.1.

Proof. If we transpose the matrix $\boldsymbol{A}$ we get the first companion matrix of the matrix polynomial $P(z)'$. Thus by applying (3.3)

$$\begin{aligned} H(z) &= \boldsymbol{C}\left(z\boldsymbol{I}_N - \boldsymbol{A}\right)\boldsymbol{B} = \boldsymbol{C}\left(z\boldsymbol{I}_N - \boldsymbol{A}'\right)'\boldsymbol{B} \\ &= \left(P(z)^{-1}, ..., z^{p-1}P(z)^{-1}\right)\boldsymbol{B} \\ &= P(z)^{-1}Q(z). \end{aligned}$$

Hence the transfer function of the MCARMA process and the state space representation are equal. □

The next representation is the proposed representation of Marquardt and Stelzer in [24].

Lemma 3.7 (Observer canonical representation)**.** *Let $\{\boldsymbol{Y}_t\}$ be a MCARMA process as in Definition 3.1. Then a clss of this process is given by $(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C})$, where $\boldsymbol{A}$ is the first companion matrix of the matrix polynomial $P(z)$, $\boldsymbol{B} = \left(\beta_1', ..., \beta_p'\right)'$ and $\boldsymbol{C} = (\boldsymbol{I}_n, \boldsymbol{0}, ..., \boldsymbol{0})$. The betas appearing in the definition of $\boldsymbol{B}$ are given recursively by*

$$\beta_{p-j} := -\sum_{i=1}^{q-j} A_i\beta_{p-j-i} + B_{q-j},\ j \in \{0, ..., q\}$$

and $\beta_j = 0$ for $j \in \{1, ..., p-q-1\}$.

Proof. We will prove that this form is just a transformation of the representation in Lemma 3.6. It can be easily seen that if T is a non singular matrix, then the transfer function of the process can also be represented via

$$H(z) = \boldsymbol{C}\left(z\boldsymbol{I}_N - \boldsymbol{A}\right)\boldsymbol{B} = \boldsymbol{C}T\left(z\boldsymbol{I}_N - T^{-1}\boldsymbol{A}T\right)T^{-1}\boldsymbol{B}.$$

As a result the first step is to find a matrix T which transforms the second companion matrix into the first companion matrix. It can be checked quite easily that

$$T = \begin{pmatrix} A_{p-1} & A_{p-2} & ... & A_1 & \boldsymbol{I}_n \\ A_{p-2} & & \iddots & \iddots & \\ \vdots & \iddots & \iddots & & \\ A_1 & \iddots & & 0 & \\ \boldsymbol{I}_n & & & & \end{pmatrix}.$$

Once again the inversion of this matrix can be done with the help of the Gaussian elimination, which yields

$$T^{-1} = \begin{pmatrix} & & & & \boldsymbol{I}_n \\ & 0 & & \iddots & T_1 \\ & & \iddots & \iddots & \vdots \\ & \iddots & \iddots & & T_{p-2} \\ \boldsymbol{I}_n & T_1 & ... & T_{p-2} & T_{p-1} \end{pmatrix},$$

where $T_k = -\sum_{i=1}^{k} A_i T_{k-i}, k \in \{1, ..., p-1\}$ and where we set $T_0 = \boldsymbol{I}_n$. Applying this transformation to $\boldsymbol{C}$ and $\boldsymbol{B}$ from the representation in Lemma 3.6 gives $\boldsymbol{C}T = (\boldsymbol{0}, ...\boldsymbol{0}, \boldsymbol{I}_n)$ and

$$T^{-1}B = \left(\beta_1^{*\prime}, ..., \beta_p^{*\prime}\right)',$$

where $\beta_j^* = 0, j = 1, ..., p-q-1$ and $\beta_{p-j}^* = \sum_{i=1}^{q-j} T_i B_{q-j-i} + B_{q-j}$.
Finally we have to show that $\beta_{p-j}^* = \beta_{p-j}, j = 0, ..., q$. We will do this by induction.
Since $\beta_{p-q}^* = B_0 = \beta_{p-q}$ and $\beta_{p-q+1}^* = -A_1 B_0 + B_1 = \beta_{p-q+1}$ we have shown that the relation holds for $j = q$ and $j = q-1$. Therefore assume that it also holds for some $j = q-k$, $k \in \{2, ..., q-1\}$. Then

$$\begin{aligned}
\beta_{p-q+k+1} &= -\sum_{i=1}^{k+1} \beta_{p-q+k+1-i} + B_{k+1} \\
&\overset{IH}{=} -\sum_{i=1}^{k+1} A_i \left(\sum_{j=1}^{k+1-i} T_j B_{k+1-i-j} + B_{k+1-i} \right) + B_{k+1} \\
&= -\sum_{i=1}^{k+1} A_i \left(\sum_{j=0}^{k+1-i} T_j B_{k+1-i-j} \right) + B_{k+1} \\
&\overset{l=i+j}{=} -\sum_{i=1}^{k+1} A_i \left(\sum_{l=i}^{k+1} T_{l-i} B_{k+1-l} \right) + B_{k+1} \\
&= -\sum_{l=1}^{k+1} \left(\sum_{i=1}^{l} A_i T_{l-i} \right) B_{k+1-l} + B_{k+1} \\
&= \sum_{l=1}^{k+1} T_l B_{k+1-l} + B_{k+1} \\
&= \beta_{p-q+k+q}^*.
\end{aligned}$$

□

The last representation that we will present in this thesis will always result in a minimal state space representation. Since the proof involves many new definitions and technicalities which does not give us a deeper inside, we will omit them here. They can be found in detail in [20].

Lemma 3.8 (Echelon realization)**.** *Let $H(z)$ be a transfer function of a n-dimensional MCARMA process with given Kronecker indices $(\nu_1, ..., \nu_n)$. Then a unique minimal state space realization is given by the following triplet.*
$\boldsymbol{A} = (A_{ij})_{i,j=1,...,n}$, *where*

$$A_{ii} = \begin{pmatrix} 0 & 1 & 0 & ... & 0 \\ 0 & 0 & 1 & & 0 \\ \vdots & \vdots & & \ddots & \vdots \\ 0 & 0 & 0 & ... & 1 \\ \alpha_{ii,1} & \alpha_{ii,2} & \alpha_{ii,3} & ... & \alpha_{ii,\nu_i} \end{pmatrix} \in \mathbb{R}^{\nu_i \times \nu_i}, A_{ij} = \begin{pmatrix} 0 & ... & ... & ... & .. & 0 \\ \vdots & & & & & \vdots \\ 0 & ... & ... & ... & ... & 0 \\ \alpha_{ij,1} & ... & \alpha_{ij,\nu_{ij}} & 0 & ... & 0 \end{pmatrix} \in \mathbb{R}^{\nu_i \times \nu_j},$$

and $\nu_{ij} := \min(\nu_i + \mathbf{1}_{\{i>j\}}, \nu_j)$,

$\boldsymbol{B} = (b_{ij}) \in \mathbb{R}^{N\times d}$ *is unrestricted,*

$$\boldsymbol{C} = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 & 1 & 0 & \dots & 0 & 0 & 0 & \dots & 0 \\ \vdots & & & & & & & & & & & & \vdots \\ 0 & 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & 1 & 0 & \dots & 0 \end{pmatrix} \in \mathbb{R}^{n\times N},$$

whereas the ones are in the columns $1, (\nu_1+1), \dots, (\nu_1+\dots+\nu_{n-1}+1)$.
Furthermore a unique left matrix fraction description $P(z)^{-1}Q(z)$ *of* $H(z)$ *is given by*

$$(P(z))_{ij} = \mathbf{1}_{\{i=j\}} z^{\nu_i} - \sum_{k=1}^{\nu_{ij}} \alpha_{ij,k} z^{k-1} \text{ and } (Q(z))_{ij} = \sum_{k=1}^{\nu_i} \kappa_{\nu_1+\dots+\nu_{i-1}+k,j} z^{k-1},$$

and the coefficient $\kappa_{i,j}$ *is the* (i,j)*th entry of the matrix* $K = M\boldsymbol{B}$, *where* $M = (M_{ij})_{i,j=1,\dots,n} \in \mathbb{R}^{N\times N}$ *is given by*

$$M_{ii} = \begin{pmatrix} -\alpha_{ii,2} & -\alpha_{ii,3} & \dots & -\alpha_{ii,\nu_i} & 1 \\ -\alpha_{ii,3} & -\alpha_{ii,4} & \dots & 1 & \\ \vdots & & \iddots & & \\ -\alpha_{ii,\nu_i} & 1 & & & \\ 1 & & & 0 & \end{pmatrix} \in \mathbb{R}^{\nu_i\times\nu_i}, M_{ij} = \begin{pmatrix} -\alpha_{ij,2} & -\alpha_{ij,3} & \dots & -\alpha_{ij,\nu_{ij}} & 0 \\ -\alpha_{ij,3} & -\alpha_{ij,4} & \dots & 0 & 0 \\ \vdots & & \iddots & \vdots & \vdots \\ -\alpha_{ij,\nu_{ij}} & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 \\ \vdots & & & & \vdots \\ 0 & 0 & \dots & 0 & 0 \end{pmatrix} \in \mathbb{R}^{\nu_i\times\nu_j}.$$

The biggest advantage of this representation is of course that it is minimal. The simplest way to see that this representation is minimal, is that $\boldsymbol{A}$ is a $\sum_{i=1}^n \nu_i \times \sum_{i=1}^n \nu_i$ matrix, but we know from Theorem 3.4 that the sum over the Kronecker indices is equal to the McMillan degree, and hence $(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C})$ must be minimal. Another advantage in contrast to all the other representations in this thesis is, that the leading coefficient of the matrix polynomial $P(z)$ do not have to be the identity matrix. Therefore it makes sense to classify the above representation not in terms of its matrix polynomial degrees p and q, but rather in terms of its Kronecker indices.

Example 3.9 (MCARMA(2,2) process)**.** Suppose that we have a 2-dimensional MCARMA process, whose transfer function has Kronecker indices $(\nu_1, \nu_2) = (2, 2)$, and which has a 2-dimensional background driving Lévy noise. Then the matrices $\boldsymbol{A}$ and $\boldsymbol{C}$ of the echelon representation have the form

$$\boldsymbol{A} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ \alpha_{11,1} & \alpha_{11,2} & \alpha_{12,1} & \alpha_{12,2} \\ 0 & 0 & 0 & 1 \\ \alpha_{21,1} & \alpha_{21,2} & \alpha_{22,1} & \alpha_{22,2} \end{pmatrix} \text{ and } \boldsymbol{C} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

Furthermore we have that

$$M = \begin{pmatrix} -\alpha_{11,2} & 1 & -\alpha_{12,2} & 0 \\ 1 & 0 & 0 & 0 \\ -\alpha_{21,2} & 0 & -\alpha_{22,2} & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \text{ and } M^{-1} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & \alpha_{11,2} & 0 & \alpha_{12,2} \\ 0 & 0 & 0 & 1 \\ 0 & \alpha_{21,2} & 1 & \alpha_{22,2} \end{pmatrix}.$$

Hence we can write $\boldsymbol{B}$ as

$$\boldsymbol{B} = M^{-1}K = \begin{pmatrix} \kappa_{2,1} & \kappa_{2,2} \\ \kappa_{1,1} + \alpha_{11,2}\kappa_{2,1} + \alpha_{12,2}\kappa_{4,1} & \kappa_{1,2} + \alpha_{11,2}\kappa_{2,2} + \alpha_{12,2}\kappa_{4,2} \\ \kappa_{4,1} & \kappa_{4,2} \\ \kappa_{3,1} + \alpha_{21,2}\kappa_{2,1} + \alpha_{22,2}\kappa_{4,1} & \kappa_{3,2} + \alpha_{21,2}\kappa_{2,2} + \alpha_{22,2}\kappa_{4,2} \end{pmatrix}.$$

Additionally we can recover the AR and MA polynomials, i.e.

$$P(z) = \boldsymbol{I}_2 z^2 - \begin{pmatrix} \alpha_{11,2} & \alpha_{12,2} \\ \alpha_{21,2} & \alpha_{22,2} \end{pmatrix} z - \begin{pmatrix} \alpha_{11,1} & \alpha_{12,1} \\ \alpha_{21,1} & \alpha_{22,1} \end{pmatrix} =: \boldsymbol{I}_2 z^2 - A_2 z - A_1,$$

$$Q(z) = \begin{pmatrix} \kappa_{2,1} & \kappa_{2,2} \\ \kappa_{4,1} & \kappa_{4,2} \end{pmatrix} z + \begin{pmatrix} \kappa_{1,1} & \kappa_{1,2} \\ \kappa_{3,1} & \kappa_{3,2} \end{pmatrix} =: K_0 z + K_1.$$

As a result we have found a parametric model for the MCARMA(2,2) process in echelon form. Before we end this example we want to see the connection between this representation and the observer canonical form. Therefore we will apply a simple transformation which changes two rows and columns.
Let

$$T := \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

then the transformed state space representation $(T^{-1}\boldsymbol{A}T, T^{-1}\boldsymbol{B}, \boldsymbol{C}T)$ is given by

$$T^{-1}\boldsymbol{A} = \begin{pmatrix} \boldsymbol{0}_{2\times 2} & \boldsymbol{I}_2 \\ -A_2 & -A_1 \end{pmatrix}, \boldsymbol{C}T = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} \text{ and}$$

$$T^{-1}\boldsymbol{B} = \begin{pmatrix} \kappa_{2,1} & \kappa_{2,2} \\ \kappa_{4,1} & \kappa_{4,2} \\ \kappa_{1,1} + \alpha_{11,2}\kappa_{2,1} + \alpha_{12,2}\kappa_{4,1} & \kappa_{1,2} + \alpha_{11,2}\kappa_{2,2} + \alpha_{12,2}\kappa_{4,2} \\ \kappa_{3,1} + \alpha_{21,2}\kappa_{2,1} + \alpha_{22,2}\kappa_{4,1} & \kappa_{3,2} + \alpha_{21,2}\kappa_{2,2} + \alpha_{22,2}\kappa_{4,2} \end{pmatrix} = \begin{pmatrix} K_1 \\ K_0 - A_1 K_1 \end{pmatrix}.$$

But this is nothing else then the observer canonical form and since the dimension of $\boldsymbol{A}$ has not changed, the observer canonical form is also minimal for this example.

3.2 Estimation of the model parameters

3.3 The quasi maximum likelihood estimation

In the first part of this section we will introduce our observation scheme. Afterwards we will show how to estimate the parameters of the MCARMA polynomials $P(z)$ and $Q(z)$ based on a quasi maximum likelihood procedure, which was proposed by [30] for the case of MCARMA processes with a Lévy noise possessing a first and second moment. Although our Lévy noise will just have a first moment, we will stick to this estimation procedure, since the quasi maximum likelihood estimator (QMLE) performs in the infinite variance equally well or even better than in the Gaussian case for univariate ARMA processes, see e.g. [13]. The last part will be the recovery of the background Lévy noise as introduced in [10].
Throughout the rest of this thesis we will assume that we observe the MCARMA process at equidistant points in time, that is we observe the output process $\boldsymbol{Y}_t$ at times $0 = t_0 < t_1 < \ldots < t_m$, $m \in \mathbb{N}$ and $t_i = i\Delta t$.

Lemma 3.10. *For the continuous-time linear state space model*

$$\mathrm{d}\boldsymbol{X}_t = \boldsymbol{A}\boldsymbol{X}_t\,\mathrm{d}t + \boldsymbol{B}\,\mathrm{d}\boldsymbol{L}_t, \qquad \boldsymbol{Y}_t = \boldsymbol{C}\boldsymbol{X}_t \tag{3.7}$$

the associated discretely observed process $\left\{\boldsymbol{Y}^{\Delta t}\right\}$ *is given by*

$$\boldsymbol{X}_k = e^{\boldsymbol{A}\Delta t}\boldsymbol{X}_{k-1} + N_k^{\Delta t}, \qquad \boldsymbol{Y}_k^{\Delta t} = \boldsymbol{C}\boldsymbol{X}_k,$$

where

$$N_k^{\Delta t} = \int_{(k-1)\Delta t}^{k\Delta t} e^{\boldsymbol{A}(k\Delta t - u)}\boldsymbol{B}\,\mathrm{d}\boldsymbol{L}_u.$$

Furthermore if the real parts of the spectrum of $\boldsymbol{A}$ *are strictly negative and if* $\boldsymbol{L}$ *has a finite second moment denoted by* $\Sigma^{\boldsymbol{L}}$ *and is centered, then* $N^{\Delta t}$ *is also centered and its covariance matrix is explicitly given by*

$$\Sigma^{\Delta t} = \int_0^{\Delta t} e^{\boldsymbol{A}u}\boldsymbol{B}\Sigma^{\boldsymbol{L}}\boldsymbol{B}'e^{\boldsymbol{A}'u}\,\mathrm{d}u.$$

Finally the autocovariance function of the discretized process for each $i \in \mathbb{N}$ *is given by*

$$\gamma_{\boldsymbol{Y}^{\Delta t}}(h) = \mathrm{Cov}\left(\boldsymbol{Y}_{i+h\Delta t}^{\Delta t}, \boldsymbol{Y}_i^{\Delta t}\right) = \boldsymbol{C}e^{\boldsymbol{A}h\Delta t}\overline{\Sigma}\boldsymbol{C}',$$

where

$$\overline{\Sigma} = \int_0^{\infty} e^{\boldsymbol{A}u}\boldsymbol{B}\Sigma^{\boldsymbol{L}}\boldsymbol{B}'e^{\boldsymbol{A}'u}\,\mathrm{d}u.$$

Proof. The solution of the SDE (3.7) is

$$\boldsymbol{X}_t = e^{\boldsymbol{A}(t-s)}\boldsymbol{X}_s + \int_s^t e^{\boldsymbol{A}(t-u)}\boldsymbol{B}\,\mathrm{d}\boldsymbol{L}_u. \tag{3.8}$$

It can be obtained in the usual way via Itô's formula for semimartingales, for details we refer to the next section. Replacing t and s by $k\Delta t$ and $(k-1)\Delta t$, resp., yields the first statement.

For the second part of the Lemma, we use Lemma 2.2. Since $\boldsymbol{L}$ has zero mean, the same immediately follows for the noise of the discretized process. Due to the stationary increments of the Lévy process $\boldsymbol{L}$, we get

$$\begin{aligned}\operatorname{Cov}\left(N_k^{\Delta t}\right) &= \operatorname{Cov}\left(\int_0^{\Delta t} e^{\boldsymbol{A}u}\boldsymbol{B}\,\mathrm{d}\,\boldsymbol{L}_u\right)\\ &\overset{\text{Lemma 2.2}}{=} \int_0^{\Delta t} e^{\boldsymbol{A}u}\boldsymbol{B}\Sigma^{\boldsymbol{L}}\boldsymbol{B}'e^{\boldsymbol{A}'u}\,\mathrm{d}\,u.\end{aligned}$$

Since the real parts of the eigenvalues of $\boldsymbol{A}$ are strictly negative the first summand of the left handside of (3.8) vanishes as $s \to -\infty$. In addition with the stationary increments of the Lévy process we thus get that

$$\boldsymbol{X}_t \overset{d}{=} \int_0^{\infty} e^{\boldsymbol{A}u}\boldsymbol{B}\,\mathrm{d}\,\boldsymbol{L}_u.$$

Using Lemma 2.2 once again we obtain that

$$\operatorname{Cov}(\boldsymbol{X}_t) = \overline{\Sigma}.$$

Furthermore by replacing t and s by $t+u$ and t, resp. in (3.8) we obtain

$$X_{t+v} = e^{\boldsymbol{A}v}\boldsymbol{X}_t + \int_t^{t+v} e^{\boldsymbol{A}(t+h-u)}\boldsymbol{B}\,\mathrm{d}\,\boldsymbol{L}_u$$

and hence by the independent increments of the Lévy process that

$$\operatorname{Cov}(\boldsymbol{X}_{t+v}, \boldsymbol{X}_t) = e^{\boldsymbol{A}v}\operatorname{Cov}(\boldsymbol{X}_t) + \operatorname{Cov}\left(\int_{-\infty}^{t} e^{\boldsymbol{A}(t-u)}\boldsymbol{B}\,\mathrm{d}\,\boldsymbol{L}_u, \int_t^{t+v} e^{\boldsymbol{A}(t+h-u)}\boldsymbol{B}\,\mathrm{d}\,\boldsymbol{L}_u\right) = e^{\boldsymbol{A}v}\overline{\Sigma}.$$

The more general result for the discretized output process $\boldsymbol{Y}^{\Delta t}$ follows by the observation equation and the replacement of t and v by i and $h\Delta t$, resp. □

Theorem 3.11 (Estimation of the MCARMA parameters [30])**.** *For a parametrized continuous-time state space model $(\boldsymbol{A}_\vartheta, \boldsymbol{B}_\vartheta, \boldsymbol{C}_\vartheta)$, where $\vartheta \in \Theta \subset \mathbb{R}^r$, and M discrete observations $y^M = \left\{y_i^{\Delta t}\right\}_{i=1,\dots,M}$ from the output process $Y_{\vartheta_0}^{\Delta t}$ sampled at equidistant points in time with mesh size Δt a quasi maximum likelihood estimator for the true parameters ϑ_0 of the MCARMA model is given by minimizing the negative log likelihood*

$$\widehat{\mathcal{L}}^{\Delta t}(\vartheta, y^M) = \sum_{k=1}^{m}\left(d\log 2\pi + \log\det V_\vartheta + \hat{\varepsilon}'_{\vartheta,k}V_\vartheta^{-1}\hat{\varepsilon}_{\vartheta,k}\right). \tag{3.9}$$

Here $\hat{\varepsilon}_{\vartheta,k}$ are the pseudo-innovations of the observed process $\boldsymbol{Y}_{\vartheta_0}^{\Delta t}$, which can be calculated recursively via

$$\widehat{\boldsymbol{X}}_{\vartheta,k} = \left(e^{A_\vartheta\Delta t} - K_\vartheta\boldsymbol{C}_\vartheta\right)\widehat{\boldsymbol{X}}_{\vartheta,k-1} + K_\vartheta y_{k-1}^{\Delta t}, \tag{3.10}$$

and

$$\hat{\varepsilon}_{\vartheta,k} = y_k^{\Delta t} - \boldsymbol{C}\widehat{\boldsymbol{X}}_{\vartheta,k}. \tag{3.11}$$

Furthermore the steady-state Kalman gain matrices K_ϑ and the pseudo-covariances V_ϑ are given by

$$K_\vartheta = \left(e^{\boldsymbol{A}_\vartheta \Delta t}\Omega_\vartheta \boldsymbol{C}_\vartheta'\right)\left(\boldsymbol{C}_\vartheta \Omega_\vartheta \boldsymbol{C}_\vartheta'\right)^{-1}, \qquad V_\vartheta = \boldsymbol{C}_\vartheta \Omega_\vartheta \boldsymbol{C}_\vartheta',$$

where the error covariances Ω_ϑ are the unique positive definite solution to the discrete-time algebraic Riccati equation

$$\Omega_\vartheta = e^{\boldsymbol{A}_\vartheta \Delta t}\Omega_\vartheta e^{\boldsymbol{A}_\vartheta' \Delta t} + \Sigma_\vartheta^{\Delta t} - \left(e^{\boldsymbol{A}_\vartheta \Delta t}\Omega_\vartheta \boldsymbol{C}_\vartheta'\right)\left(\boldsymbol{C}_\vartheta \Omega_\vartheta \boldsymbol{C}_\vartheta'\right)^{-1}\left(e^{\boldsymbol{A}_\vartheta \Delta t}\Omega_\vartheta \boldsymbol{C}_\vartheta'\right)'.$$

Finally $\Sigma^{\Delta t}$ is the covariance matrix of the Lévy noise as defined in the last Lemma 3.10. Moreover if the following assumptions are fulfilled for each $\vartheta \in \Theta$

1. *the real parts of the eigenvalues of $\boldsymbol{A}_\vartheta$ are strictly negative*
2. *the absolute values of the imaginary parts of the eigenvalues of $\boldsymbol{A}_\vartheta$ are smaller than $\pi/\Delta t$*
3. *the state space model $(\boldsymbol{A}_\vartheta, \boldsymbol{B}_\vartheta, \boldsymbol{C}_\vartheta)$ is minimal with McMillan degree N*
4. *the matrix $\boldsymbol{C}_\vartheta$ has full rank*
5. *the Lévy noise $\boldsymbol{L}_\vartheta$ is centered*
6. *the parameter space Θ is a compact subset of $\mathbb{R}^r$*
7. *the functions $\vartheta \mapsto \boldsymbol{A}_\vartheta, \vartheta \mapsto \boldsymbol{B}_\vartheta, \vartheta \mapsto \boldsymbol{C}_\vartheta$ and $\vartheta \mapsto \Sigma_\vartheta^{\boldsymbol{L}}$ are continuous*
8. *$E\|\boldsymbol{L}_\vartheta(1)\|^2 < \infty$ and $\Sigma_\vartheta^{\boldsymbol{L}}$ is non-singular*
9. *for every $\vartheta_1, \vartheta_2 \in \Theta$ the spectral densities of the output processes $\boldsymbol{Y}_{\vartheta_1}$ and $\boldsymbol{Y}_{\vartheta_2}$ are different*

then the QMLE is strongly consistent, i.e.

$$\operatorname*{argmin}_{\vartheta \in \Theta} \widehat{\mathcal{L}}^{\Delta t}(\vartheta, y^M) \overset{a.s.}{\to} \vartheta_0, \qquad M \to \infty.$$

Remark 3.12. Note that in our case with $\boldsymbol{\alpha}$-copula-stable Lévy noise, where $\boldsymbol{\alpha} \in (1,2)^d$, we neither have a finite second moment nor a spectral density and hence the last two assumptions are not fulfilled. Nevertheless we stick to the QML estimation procedure as mentioned earlier. Furthermore, since we do not have a finite covariance matrix $\Sigma^{\boldsymbol{L}}$, we will replace it formally by its empirical counterpart, which is always finite. As in the finite variance case, we do neither know the covariance matrix nor the empirical covariance matrix of the Lévy process a priori. Therefore we will include additional parameters in the estimation procedure for the (empirical) covariance matrix and we have to ensure in each run of the optimization that this matrix is positive semidefinite, which can be formulated as a constraint that the eigenvalues of $\Sigma_\vartheta^{\Delta t}$ are strictly positive.

Remark 3.13. The Assumption 1. is in my view the most important, since it ensures that the output process $\boldsymbol{Y}$ is stationary. This fact is well know for CARMA processes, but it can also be found in [24] for the multivariate case.

Remark 3.14. In the paper of Schlemm and Stelzer [30], they have chosen the echelon parametrization for the state space model, since it is always minimal (Assumption 3), the spectral densities of the output process are not identical for different estimates (Assumption 9) and the observation matrix consisting of zeros and ones has full rank (Assumption 4). In the simulation study I will also use this representation. However I will use the controller canonical form in the data analysis section, since we can use the estimation of the Lévy noise within the QML optimization without any transformations of the polynomials $P(z)$ and $Q(z)$, see below.
Furthermore unless we restrict the transfer function of the MCARMA process to a fixed value at the origin, the matrix $\boldsymbol{B}$ is unrestricted. The problem arising from this fact is that $\boldsymbol{B}$ appears in the QMLE exclusively in the determination of the covariance matrix of the discretized Lévy noise $\Sigma_\vartheta^{\Delta t}$ together with the covariance matrix $\Sigma_\vartheta^{\boldsymbol{L}}$. But take for instance the eigendecomposition of the positive semidefinite matrix $\Sigma_\vartheta^{\boldsymbol{L}} = \boldsymbol{R}_\vartheta \boldsymbol{D}_\vartheta \boldsymbol{R}_\vartheta'$, where $\boldsymbol{R}_\vartheta$ is a matrix consisting of the eigenvectors of $\Sigma_\vartheta^{\boldsymbol{L}}$ and $\boldsymbol{D}_\vartheta$ is a diagonal matrix of the eigenvalues. Then there are two possible estimates for $\widehat{\boldsymbol{B}}$ and $\widehat{\Sigma^{\boldsymbol{L}}}$

$$\begin{aligned}
\Sigma_\vartheta^{\Delta t} &= \int_0^{\Delta t} e^{\boldsymbol{A}_\vartheta u} \boldsymbol{B}_\vartheta \Sigma_\vartheta^{\boldsymbol{L}} \boldsymbol{B}_\vartheta' e^{\boldsymbol{A}_\vartheta' u} \,\mathrm{d}\, u \\
&= \int_0^{\Delta t} e^{\boldsymbol{A}_\vartheta u} \left(\boldsymbol{B}_\vartheta \boldsymbol{R}_\vartheta\right) \boldsymbol{D}_\vartheta \left(\boldsymbol{B}_\vartheta \boldsymbol{R}_\vartheta\right)' e^{\boldsymbol{A}_\vartheta' u} \,\mathrm{d}\, u \\
&= \int_0^{\Delta t} e^{\boldsymbol{A}_\vartheta u} \boldsymbol{B}_\vartheta^* \Sigma_\vartheta^{*,\boldsymbol{L}} \left(\boldsymbol{B}_\vartheta^*\right)' e^{\boldsymbol{A}_\vartheta' u} \,\mathrm{d}\, u.
\end{aligned}$$

One alternative to the above given QMLE is to use the Kalman-Lévy filter instead of the usual Kalman filter as described in [14]. The key idea is to base the filter not on the minimization of the covariance matrix as it is done in the normal Kalman filter, but on the tails of the distribution of the errors. That is the gain matrix K_ϑ in (3.10) is the matrix that minimizes the "tail covariance" of the pseudo innovations $\hat{\varepsilon}_{\vartheta,k}$ in (3.11). The tail covariance is defined as the scale factors of the distributions of the products $\hat{\varepsilon}_{\vartheta,k}^{(i)} \hat{\varepsilon}_{\vartheta,k}^{(j)}, i,j = 1,...,n$, where the scale factors for univariate distributions with a power law tail are given by $SF_\mp = \lim_{\lambda\to\infty} \lambda^\alpha P(X \lessgtr \mp\lambda)$. Explicit representations of the scale factors for univariate stable distributions are given in [29] Property 1.2.15, that is $SF_\mp = c^\mp (1-\alpha)\left(\Gamma(2-\alpha)\cos(0.5\pi\alpha)\right)^{-1}$, $\alpha \neq 1$, where $\Gamma(x)$ is the Gamma function and the stable distribution is parametrized as in Example 2.3. In the mentioned work [14] the authors assume that the errors are linear combinations of independent symmetric random variables and in this case they were able to solve the optimization problem explicitly. Whereas in our model the errors have a more complicated dependence structure for which the optimization has to be done in every single step which is not applicable even for very small dimensional problems. On the contrary for univariate problems this filter seems to be preferable, but we will not concentrate on this, since we are interested in purely multidimensional CARMA processes.

Another alternative to the QMLE might be a multivariate extension of the estimation procedure for CARMA parameters which was recently published by Fasen et al. in [17]. Here the key idea is to approximate the normalized power transfer function by the smoothed self-normalized periodogram and then the estimator is given as the usual least squares estimator. My multivariate extension of the smoothed self-normalized periodogram for M samples of a MCARMA process is given by

$$\widehat{T}_{M,\boldsymbol{Y}^{\Delta t}}(\omega) := \sum_{|j|\leq c_M} W_M(j) \frac{\sum_{l=1}^{M} y_l^{\Delta t} e^{-i\left(\omega+\frac{j}{M}\right)l} \left(\sum_{l=1}^{M} y_l^{\Delta t} e^{-i\left(\omega+\frac{j}{M}\right)l}\right)^*}{\sum_{l=1}^{M} y_l^{\Delta t} \left(y_l^{\Delta t}\right)'},$$

where $W_M : \mathbb{Z} \mapsto \mathbb{R}$ is a positive and symmetric weighting function which is normalized by $\sum_{|j|\leq c_M} W_M(j) = 1$ and for which $\max_{|j|\leq c_M} W_M^2(j) = o\left(c_M^{-1}\right)$ as $M \to \infty$. c_M is the bandwidth of the weighting function s.t. $\frac{c_M}{M\Delta t} \to 0$ as $M \to \infty$. The division in the equation above is meant componentwise and $*$ denotes the conjugate transpose operator. There are many possible choice for the weighting function and the bandwidth, e.g. the Daniell smoother $W_M(j) = (2c_M + 1)^{-1}$ and $c_M = \left\lfloor \sqrt{M\Delta t} \right\rfloor$ fulfill the above requirements.
The normalized power transfer function is given as

$$S(\omega) := \frac{P(i\omega)^{-1}Q(i\omega)\left(P(i\omega)^{-1}Q(i\omega)\right)^*}{\int_0^\infty \boldsymbol{C}e^{\boldsymbol{A}u}\boldsymbol{B}\left(\boldsymbol{C}e^{\boldsymbol{A}u}\boldsymbol{B}\right)' \mathrm{d}\, u}.$$

If the process $\boldsymbol{Y}^{\Delta t}$ is univariate and if it further holds that $\Delta t \to 0$ as $M \to \infty$ as well as that Assumption 1 in Theorem 3.11 is fulfilled than it was shown in [17] that

$$\Delta t \widehat{T}_{M,\boldsymbol{Y}^{\Delta t}}(\omega\Delta t) \xrightarrow{P} S(\omega), \qquad M \to \infty.$$

Based on this result, we can construct a spectral estimator, which minimizes the squared real and imaginary parts of the distance $D(\omega_i) = S(\omega_i) - \Delta t \widehat{T}_{M,\boldsymbol{Y}^{\Delta t}}(\omega_i \Delta t)$ for $l \in \mathbb{N}$ different frequencies ω_i, $i = 1, ..., l$, i.e.

$$\widehat{\vartheta} := \underset{\vartheta\in\Theta}{\operatorname{argmin}} \sum_{k=1}^{l} \left| \sum_{1\leq i\leq j\leq n} \Re\left((D(\omega_k))_{ij}\right) + \sum_{1\leq i<j\leq n} \Im\left((D(\omega_k))_{ij}\right) \right|^2.$$

Note that the above estimator is already constructed for the multivariate framework. I also tried to find a similar result for the multivariate case, but the problem was always the non-commutativity of the matrices arising from the state space representation. Therefore I was not able to split the smoothed self-normalized periodogram into multiplicative parts and to show their convergence separately as it was done in [17].
Furthermore Figure 3.1 displays the sample paths, the smoothed self-normalized periodograms and the normalized power transfer functions for two simulated bivariate MCARMA processes. The periodgrams and the power transfer functions are presented in terms of the single components and the real/imaginary parts of the cospectrum. We observe that for the first process for which $S(0) \approx 0$ the asymptotic relation between the periodogram and the power transfer

function seems to hold also in the multivariate setup, whereas for the second process we would see the same result if we would scale the power transfer function by a certain constant. Since I saw this relationship for various other simulated processes, it seems that it is possible to construct also a spectral estimate in the multivariate case, which needs an additional multiplicative factor for the smoothed self-normalized periodogram.

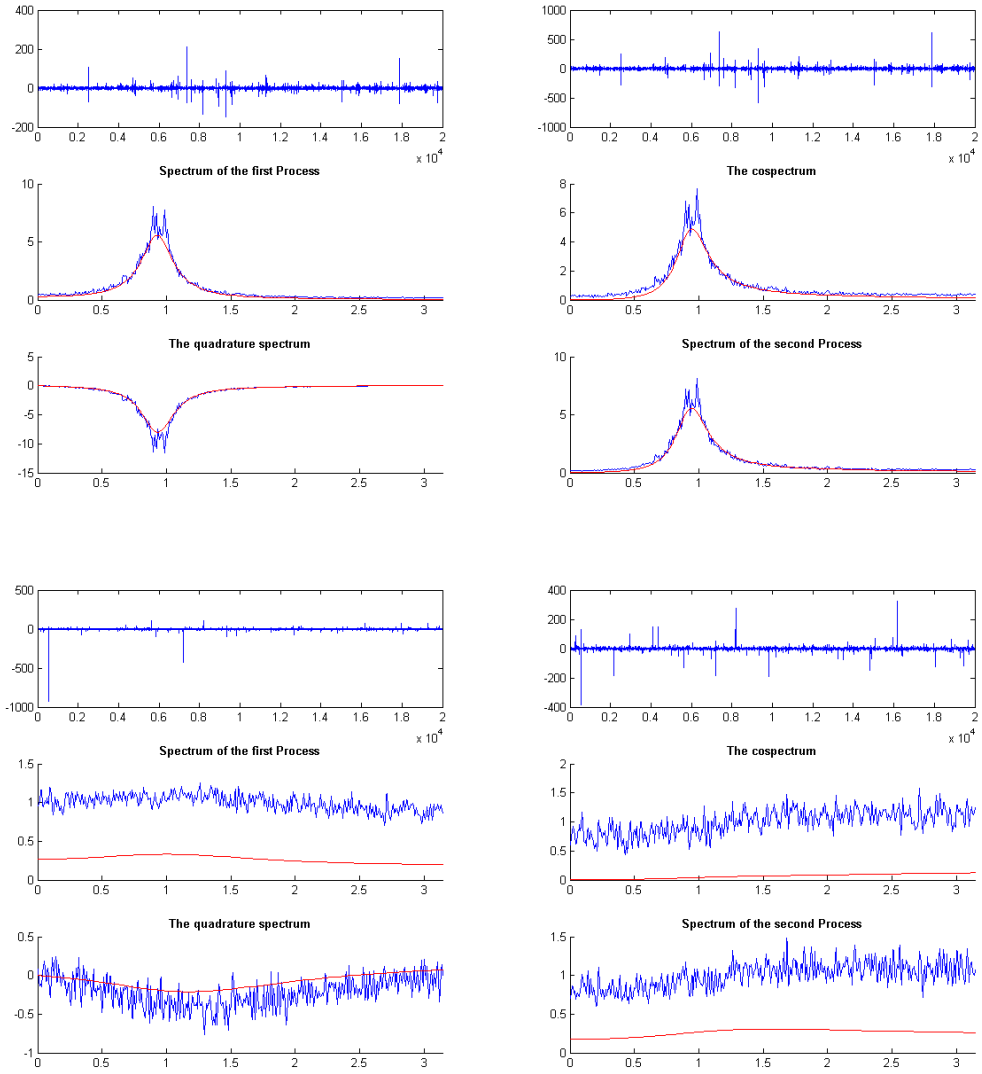


Figure 3.1: Trajectories, the smoothed self-normalized periodogram (blue) and the normalized power transfer function (red) for two simulated MCARMA processes with different parameters, using the echelon representation and a Lévy noise with two independent 1.6-stable margins.

3.4 Recovery of the Lévy increments

In the following we will assume that the state space representation of the MCARMA process is in controller canonical form. For the recovery of the Lévy increments we follow the approach of Brockwell and Schlemm given in [10]. The key idea is to "invert" the state space representation such that we can recover the state vector $\boldsymbol{X}$ from the output process $\boldsymbol{Y}$, which can than be used to obtain the Lévy increments.
First recall that the controller canonical form is given by the triple

$$\boldsymbol{A} = \begin{pmatrix} 0 & \boldsymbol{I}_d & 0 & ... & 0 \\ 0 & 0 & \boldsymbol{I}_d & ... & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & ... & \boldsymbol{I}_d \\ -A_p & -A_{p-1} & -A_{p-2} & ... & -A_1 \end{pmatrix}, \boldsymbol{B} = \begin{pmatrix} \boldsymbol{0} \\ \vdots \\ \boldsymbol{0} \\ \boldsymbol{I}_d \end{pmatrix}, \boldsymbol{C} = \begin{pmatrix} B_q' \\ \vdots \\ B_0' \\ \boldsymbol{0} \\ \vdots \\ \boldsymbol{0} \end{pmatrix}'.$$

Lemma 3.15. *If the n-dimensional MCARMA(p,q) process is in controller canonical form with d-dimensional Lévy noise and the following assumptions hold*

1. $d \leq n$
2. B_0 *and* $B_0' B_q$ *have full rank* d
3. *the roots of the polynomial* $\det B_0^{\sim 1} Q(z)$ *have strictly negative real parts, where* $B_0^{\sim 1}$ *denotes the generalized left inverse, i.e.* $B_0^{\sim 1} B_0 = \boldsymbol{I}_d$

then we can recover the state vector for the first dq components via the SDE

$$\mathrm{d}\,\boldsymbol{X}_t^{\{q\}} = \widetilde{\boldsymbol{A}} \boldsymbol{X}_t^{\{q\}}\,\mathrm{d}\,t + \widetilde{\boldsymbol{B}} \boldsymbol{Y}_t\,\mathrm{d}\,t, \tag{3.12}$$

where

$$\boldsymbol{X}_t^{\{q\}} := \left(\left(\boldsymbol{X}_t^{(1)}\right)', ..., \left(\boldsymbol{X}_t^{(q)}\right)'\right)' = \left(\left(X_{1,t}, ..., X_{d,t}\right), ..., \left(X_{(q-1)d+1,t}, ..., X_{qd,t}\right)\right)'$$

is the upper q-block-truncation of $\boldsymbol{X}_t$*,*

$$\widetilde{\boldsymbol{A}} = \begin{pmatrix} 0 & \boldsymbol{I}_d & 0 & ... & 0 \\ 0 & 0 & \boldsymbol{I}_d & ... & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & ... & \boldsymbol{I}_d \\ -B_0^{\sim 1} B_q & -B_0^{\sim 1} B_{q-1} & -B_0^{\sim 1} B_{q-2} & ... & -B_0^{\sim 1} B_1 \end{pmatrix} \text{ and } \widetilde{\boldsymbol{B}} = \begin{pmatrix} \boldsymbol{0} \\ \vdots \\ \boldsymbol{0} \\ B_0^{\sim 1} \end{pmatrix}. \tag{3.13}$$

This SDE can be solved in the usual way. Thus

$$\boldsymbol{X}_t^{\{q\}} = e^{\widetilde{\boldsymbol{A}}(t-s)} \boldsymbol{X}_s^{\{q\}} + \int_s^t e^{\widetilde{\boldsymbol{A}}(t-u)} \widetilde{\boldsymbol{B}} \boldsymbol{Y}_u\,\mathrm{d}\,u. \tag{3.14}$$

Furthermore the remaining $p-q-1$ q-blocks $\boldsymbol{X}^{(j)}$, $j=q+1,...,p$ are then given by

$$\boldsymbol{X}_t^{(j)} = \boldsymbol{E}_q \left(\widetilde{\boldsymbol{A}}^{j-q} \boldsymbol{X}_t^{\{q\}} + \sum_{k=0}^{j-q-1} \widetilde{\boldsymbol{A}}^{j-q-k-1} \widetilde{\boldsymbol{B}} D \boldsymbol{Y}_t \right),$$

where $\boldsymbol{E}_q := (\boldsymbol{0},...,\boldsymbol{0},\boldsymbol{I}_d)$ is the matrix analogue of a raw unit vector and D denotes the differential operator.

Proof. On the one hand, because of the companion structure of $\boldsymbol{A}$ in the canonical controller form, we get for each d-block $\boldsymbol{X}_t^{(j)}$, $j=1,...,q$ of the state vector $\boldsymbol{X}_t$ that $\mathrm{d}\,\boldsymbol{X}_t^{(j)} = \boldsymbol{X}_t^{(j+1)}\,\mathrm{d}\,t$. On the other hand, if we insert the observation equation $\boldsymbol{Y}_t = \boldsymbol{C}\boldsymbol{X}_t$ into (3.12) and calculate each d-block separately, then we obtain the same result, which shows the first statement. The result for the remaining part of $\boldsymbol{X}_t$ can be calculated easily via induction. □

After expressing the state space vector in terms of the output process we are able to recover the Lévy noise. For this purpose we obtain from the last d raws of the state equation of the canonical controller form that

$$\mathrm{d}\,\boldsymbol{X}_t^{(p)} = -\sum_{k=1}^{p} A_{p+1-k} \boldsymbol{X}_t^{(k)}\,\mathrm{d}\,t + \mathrm{d}\,\boldsymbol{L}_t,$$

which can be rearranged after integrating from 0 to t as

$$\boldsymbol{L}_t = \boldsymbol{X}_t^{(p)} - \boldsymbol{X}_0^{(p)} + \int_0^t \sum_{k=1}^{p} A_{p+1-k} \boldsymbol{X}_t^{(k)}\,\mathrm{d}\,t.$$

Using this representation of the Lévy noise in conjunction with the knowledge of the state space vectors from the prior lemma, we obtain the following theorem.

Theorem 3.16 (Recovery of the Lévy increments, [10])**.** *In the framework of Lemma 3.15, the Lévy increments $\Delta\boldsymbol{L}_k := \boldsymbol{L}_k - \boldsymbol{L}_{k-1}$ are given by*

$$\begin{aligned}
\Delta\boldsymbol{L}_k \;=\; & \sum_{j=0}^{p-q-1} \left(\boldsymbol{E}_q \widetilde{\boldsymbol{A}}^{p-q-1-j} \widetilde{\boldsymbol{B}} + \sum_{l=0}^{p-q-2} A_{p-q-l-1} \boldsymbol{E}_q \widetilde{\boldsymbol{A}}^{l-j} \widetilde{\boldsymbol{B}} \right) \left(D^j \boldsymbol{Y}_k - D^j \boldsymbol{Y}_{j-1} \right) \\
& + \left(\underline{A}_q \widetilde{\boldsymbol{A}}^{-1} + \sum_{l=1}^{p-q} A_{p-q-k+1} \boldsymbol{E}_q \widetilde{\boldsymbol{A}}^{l-1} + \boldsymbol{E}_q \widetilde{\boldsymbol{A}}^{p-q} \right) \left(\boldsymbol{X}_k^{\{q\}} - \boldsymbol{X}_{k-1}^{\{q\}} \right) \\
& + A_p \left(\tilde{B_0^{-1}} B_q \right)^{-1} \tilde{B_0^{-1}} \int_{k-1}^{k} \boldsymbol{Y}_u\,\mathrm{d}\,u,
\end{aligned}$$

where is $\boldsymbol{X}_k^{\{q\}}$ given in (3.14) and $\underline{A}_q := [A_p,...,A_1]$.

Remark 3.17. Note that Theorem 3.16 is only suited for processes which are observed continuously, since otherwise we are not able to integrate or differentiate $\boldsymbol{Y}$. The simplest solution to correct this problem is to use the discrete approximations, namely a trapezoidal approximation for the integrals as well as difference quotients for the differentials.

Since the parameter estimation of the MCARMA process took to much time, I was not able to construct an estimation procedure for the Lévy copula parameters. A procedure which comes closest to the estimation of the Lévy copula parameters is the one proposed by Esmaeili et al. in [15]. There the estimation of the parameters is split up into two parts. Firstly the estimation of the two marginals in terms of a maximum likelihood estimation of a truncated compound Poisson process. And secondly the estimation of the Lévy copula parameters also in terms of a MLE, but know on the the truncated bivariate compound Poisson process. Although in the same paper the authors derived an explicit representation for the MLEs for a Lévy process with different stable margins and a Clayton copula, this method is not applicable in our case. This is due to the fact that they derived the MLE's only for a positive Clayton Lévy copula. But this is essentially no restriction, since we get the general result by treating each quadrant separately as in the following Lemma 3.18. We also calculated the corresponding expressions for the bidirectional Lévy copula with stable margins, but we will not state them here since the terms are much longer than the ones in Lemma 3.18. The by far more important restriction is that the observation scheme in [15] can detect, when a jump occurs, whereas in our framework we just see the aggregated jumps of a fixed interval with length Δt. At a first glance my proposal to solve this problem is to derive the likelihood of the $\lambda^{(\varepsilon)} := \overline{\nu}^{(\varepsilon)}(\varepsilon,\varepsilon)$-th convolution of the truncated bivariate compound Poisson distribution, where $\lambda^{(\varepsilon)}$ is the jump intensity and $\overline{\nu}^{(\varepsilon)}$ is the tail integral of the truncated compound Poisson process.

Lemma 3.18. *Let $\boldsymbol{L}_t$ be a bivariate Lévy process with unbounded Lévy measure ν concentrated on $\mathbb{R}^2\backslash\{0\}$ and the Lévy copula C, which is different from the independence Lévy copula. Consider only those jumps, which are larger in absolute terms than ε in both component processes. Then the Lévy copula of the resulting compound Poisson process is given by*

$$\widetilde{C}^{(\varepsilon)}(u,v) = C\left(C_1^{\leftarrow}\left(\operatorname{sgn}(v)u, \lambda^{(\operatorname{sgn}(v)\varepsilon)}\right), C_2^{\leftarrow}\left(\lambda^{(\operatorname{sgn}(u)\varepsilon)}, \operatorname{sgn}(u)v\right)\right),$$

where $u \in \left[\operatorname{sgn}(v)\lambda^{(-\varepsilon,\operatorname{sgn}(v)\varepsilon)}, \operatorname{sgn}(v)\lambda^{(\varepsilon,\operatorname{sgn}(v)\varepsilon)}\right]\backslash\{0\}$,
$v \in \left[\operatorname{sgn}(u)\lambda^{(\operatorname{sgn}(u)\varepsilon,-\varepsilon)}, \operatorname{sgn}(u)\lambda^{(\operatorname{sgn}(u)\varepsilon,\varepsilon)}\right]\backslash\{0\}$,
$C_k^{\leftarrow}$, $k = 1,2$ *is the inverse of* C *w.r.t. to the* k*-th argument,*
$\lambda^{(\operatorname{sgn}(u)\varepsilon,\operatorname{sgn}(v)\varepsilon)} = \overline{\nu}(\operatorname{sgn}(u)\varepsilon, \operatorname{sgn}(v)\varepsilon)$,
$\lambda_1^{(\operatorname{sgn}(u)\varepsilon)} = \overline{\nu}_1(\operatorname{sgn}(u)\varepsilon)$ *and* $\lambda_2^{(\operatorname{sgn}(v)\varepsilon)} = \overline{\nu}_1(\operatorname{sgn}(v)\varepsilon)$.

Proof. The proof is essentially the same like in [16] Lemma 4.1, if we set the truncated marginal Lévy measures for $x, y \in [-\infty,-\varepsilon) \cup (\varepsilon,\infty]$ to

$$\overline{\nu}_1^{(\operatorname{sgn}(y)\varepsilon)} = \operatorname{sgn}(y)\overline{\nu}\left(x, \operatorname{sgn}(y)\varepsilon\right)$$

and

$$\overline{\nu}_2^{(\operatorname{sgn}(x)\varepsilon)} = \operatorname{sgn}(x)\overline{\nu}\left(\operatorname{sgn}(x)\varepsilon, y\right).$$

Here we had to set the leading signs manually in order to make the truncated marginal tail integrals consistent with the definition of the tail integral. □

4 The Model

Throughout this section we will work in the filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}\}_t, \mathcal{Q})$, which satisfies the usual hypotheses.
We will model our futures $f^0(t,T) = (f_i^0(t,T), i = 1, ..., n)$ in the HJM framework as proposed by [2], whereas we allow a tempered-copula-stable driving noise. Therefore we get

$$\mathrm{d}\, f^0(t,T) = \Sigma_0(t,T)\, \mathrm{d}\, \boldsymbol{L}_t, \tag{4.1}$$

where $\boldsymbol{L}_t$ is a d-dimensional $(\boldsymbol{\alpha}, \boldsymbol{\theta})$-tempered-copula-stable Lévy process as defined in Definition 2.8 with $\boldsymbol{\alpha} \in (1,2)^d$, $\Sigma_0(t,T) = \boldsymbol{C} e^{\boldsymbol{A}(T-t)} \boldsymbol{B} \in \mathbb{R}^{n\times d}$, $\boldsymbol{A} \in \mathbb{R}^{N\times N}$, $\boldsymbol{B} \in \mathbb{R}^{N\times d}$ and $\boldsymbol{C} \in \mathbb{R}^{n\times N}$. Here T denotes the "Time of maturity". Moreover I assume that

$$f^0(t,T) = f^*(t,T) - \Lambda^*(T),$$

where $\Lambda^*(T)$ is the deterministic part of the future without a trend capturing a possible seasonality and an intercept, which sets the current future price to zero, and $f^*(t,T)$ is the stochastic part. Note that the deterministic part depends only on the time of maturity and not on the current time, since I assume that for example a seasonal adjustment for an energy future price applies exclusively to the time when the actual delivery of the electricity takes place.
In the next step, we will transform our futures such that they do not depend on the time of maturity T, but instead on the time to maturity x, which was first proposed by [6]. We can already see that our volatility $\Sigma_0(t,T)$ depends exclusively on the time to maturity and not on the current time of the future. For many investment products this seems to be reasonable, but for commodities, like e.g. electricity, which show a clear seasonal pattern, this assumption remains questionable. Therefore we introduced the seasonality directly in the observed future price. But before we can start, we need to introduce the semigroup of the left translations.

Lemma 4.1. *For every $f \in C[0,\infty)$ we get that the left translations are given by*

$$f(t+x) = e^{\boldsymbol{F}t} f(x),$$

where $\boldsymbol{F} := \frac{\partial}{\partial x}$.

Proof. We will just proof this result for functions which have derivatives of all orders, i.e. $f \in C^\infty[0,\infty)$. This condition is sufficient for our purposes, since $\Sigma_0(t,\cdot) \in C^\infty[0,\infty)$. The more general result can be found in any standard textbook on semigroups, e.g. [21]. Writing the exponential function in its series representation, we get

$$\begin{aligned} e^{\boldsymbol{F}t} f(x) &= \sum_{n=0}^{\infty} \frac{\boldsymbol{F}^n t^n}{n!} f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(x)}{n!} (t+x-x)^n \\ &= f(t+x), \end{aligned}$$

whereas the second last equation is just a Taylor series expansion around x. □

With this result we are able to write the HJM equation in terms of the Musiela parameterization.

Lemma 4.2 (The Heath-Jarrow-Morton-Musiela Equation (HJMM))**.** *If $f(t,x) := f^0(t,t+x)$ and $\Sigma(x) := \Sigma_0(t,t+x)$, then the HJM equation (4.1) can be rewritten as*

$$\mathrm{d}\, f(t,x) = \boldsymbol{F} f(t,x)dt + \Sigma(x)\,\mathrm{d}\,\boldsymbol{L}_t. \tag{4.2}$$

Here $\boldsymbol{F}$ means differentiation w.r.t. the second argument.

Proof. From the previous lemma we know that $f(t,x) = e^{\boldsymbol{F}t} f^0(t,x)$. Therefore by applying Itô's formula for multivariate semimartingales, we obtain

$$\begin{aligned}
\int_0^t \mathrm{d}\left(e^{\boldsymbol{F}s} f^0(s,x)\right) &= \int_0^t \boldsymbol{F} e^{\boldsymbol{F}s} f^0(s,x)\,\mathrm{d}\,s + \int_0^t e^{\boldsymbol{F}s}\,\mathrm{d}\, f^0(s,x) \\
&+ \sum_{0\le s\le t} \underbrace{e^{\boldsymbol{F}s} f^0(s,x) - e^{\boldsymbol{F}s-} f^0(s-,x) - \Delta f^0(s,x) e^{\boldsymbol{F}s}}_{=0,\text{ since } f^0 \text{ is continuous in its second argument}} \\
&= \int_0^t \boldsymbol{F} f(s,x)\,\mathrm{d}\,s + \int_0^t \Sigma(x)\,\mathrm{d}\,\boldsymbol{L}_s.
\end{aligned}$$

Since (4.2) is just the shorthand hereof, we are done. □

In the next proposition we connect our HJMM equation to a finite dimensional realization.

Proposition 4.3 (Finite dimensional realization)**.** *The HJMM equation*

$$\mathrm{d}\, f(t,x) = \boldsymbol{F} f(t,x)dt + \Sigma(x)\,\mathrm{d}\,\boldsymbol{L}_t$$

has a finite dimensional realization which is given by

$$\mathrm{d}\,\boldsymbol{X}_t = \boldsymbol{A}\boldsymbol{X}_t\,\mathrm{d}\,t + \boldsymbol{B}\,\mathrm{d}\,\boldsymbol{L}_t,\ \boldsymbol{X}_0 = 0, \tag{4.3}$$

$$f(t,x) = \boldsymbol{C} e^{\boldsymbol{A}x}\boldsymbol{X}_t, \tag{4.4}$$

where $\boldsymbol{A}, \boldsymbol{B}$ and $\boldsymbol{C}$ are given as in the definition of $\Sigma(\cdot)$.

Proof. In [5] they prove it by solving a deterministic system and showing that this is equivalent to a certain input-output system. Afterwards they show that this also translates to the stochastic HJMM equation under the condition that the transfer functions are equal.
Nevertheless we will follow a more straightforward approach, by simply showing that both systems can be represented in the same way. To this end we start solving (4.2). We could obtain the solution by applying Itô's formula in the opposite direction as in the proof of Lemma 4.2, but it is easier to solve (4.1) directly and then transform it to the Musiela parametrization. Hence we obtain

$$f(t,x) = \int_0^t \Sigma(x+t-s)\,\mathrm{d}\,\boldsymbol{L}_s = \int_0^t \boldsymbol{C} e^{\boldsymbol{A}(x+t-s)}\boldsymbol{B}\,\mathrm{d}\,\boldsymbol{L}_s. \tag{4.5}$$

Our next step is to solve (4.3) and (4.4) explicitly. If we apply Itô's formula, we get

$$\mathrm{d}\left(e^{-\boldsymbol{A}t}\boldsymbol{X}_t\right) = \boldsymbol{A}e^{-\boldsymbol{A}t}\boldsymbol{X}_t\,\mathrm{d}\,t + e^{-\boldsymbol{A}t}\boldsymbol{A}\boldsymbol{X}_t\,\mathrm{d}\,t + e^{-\boldsymbol{A}t}\boldsymbol{B}\,\mathrm{d}\,\boldsymbol{X}_t.$$

Normally the first two terms on the right hand side cancel each other out, but now we have two matrices which are in general not commutative. Nevertheless it is true in this case. To see this we need to basic properties of the matrix exponential. Firstly, if T is a non singular matrix than $e^{T\boldsymbol{A}T^{-1}} = Te^{\boldsymbol{A}}T^{-1}$ and secondly, $e^{\mathrm{diag}(x_1,\ldots,x_n)} = \mathrm{diag}\left(e^{x_1},\ldots,e^{x_n}\right)$. Therefore we can commute the above matrices, since we can decompose $\boldsymbol{A}$ into its eigenrepresentation and then interchange the diagonal matrix and the exponential of the diagonal matrix. Note that this result only holds in the case of distinct eigenvalues, but we can find a similar result for Jordan matrices instead of diagonal matrices.
Furthermore, since $\boldsymbol{X}_0 = 0$, we get

$$\boldsymbol{X}_t = \int_0^t e^{\boldsymbol{A}(t-s)}\boldsymbol{B}\,\mathrm{d}\,\boldsymbol{L}_s,$$

and hence

$$f(t,x) = \boldsymbol{C}e^{\boldsymbol{A}x}\boldsymbol{X}_t = \int_0^t \boldsymbol{C}e^{\boldsymbol{A}(x+t-s)}\boldsymbol{B}\,\mathrm{d}\,\boldsymbol{L}_s.$$

And this is identical to (4.5). □

If we compare the finite dimensional realizations, we observe that by Lemma 3.3 they follow a MCARMA process if and only if

$$\boldsymbol{C}e^{\boldsymbol{A}x}\left(z\boldsymbol{I}_N - \boldsymbol{A}\right)^{-1}\boldsymbol{B} = P(z)^{-1}Q(z),$$

for certain matrix polynomials $P(z)$ and $Q(z)$. Unfortunately I could not find a representation which fulfills this equation, since there is no explicit solution for the matrix exponential. Only in the case of two independent process, that is the submatrices of the companion matrix $\boldsymbol{A}$ are diagonal matrices, I was able to derive an explicit solution in terms of the eigenvalues of $\boldsymbol{A}$. But even in this simple case, which is much better captured by two single CARMA models, the solution was very long. Furthermore the method of using the eigendecomposition to calculate the exponential of the companion matrix fails in the more general framework, since the eigenvectors are not simple functions of the eigenvalues of $\boldsymbol{A}$ any more. In this setup they are functions of the right solvents of the matrix polynomial $P(z)$ instead. Therefore the only possible way is to use one of the MCARMA parametrization for a fixed x, but before we follow this approach we have to adapt the model to the real world environment.
We have always assumed that the futures are settled on a specific single point in time, whereas in most commodity markets, especially in electricity markets, futures have a delivery period rather than a delivery time. The common approach to incorporate this to average the single point futures over the delivery period, i.e.

$$F(t,x_1,x_2) = \int_{x_1}^{x_2} w(u,x_1,x_2)f(t,u)\,\mathrm{d}\,u,$$

here x_1 and x_2 denote the beginning and the end point of the delivery period resp. In the following we will assume that the future is settled uniformly over the whole delivery

period and ignoring the time value of money, that is $w(u, x_1, x_2) = \frac{1}{x_2 - x_1}$. Therefore we can reformulate the future as

$$\begin{aligned} F(t, x_1, x_2) &= \frac{1}{x_2 - x_1} \int_{x_1}^{x_2} f(t, u) \, \mathrm{d}\, u \\ &= \boldsymbol{C} \frac{1}{x_2 - x_1} \int_{x_1}^{x_2} e^{\boldsymbol{A}u} \, \mathrm{d}\, u \boldsymbol{X}_t \\ &= \boldsymbol{C} \frac{1}{x_2 - x_1} \boldsymbol{A}^{-1} \left(e^{\boldsymbol{A}x_2} - e^{\boldsymbol{A}x_1} \right) \boldsymbol{X}_t \\ &=: \boldsymbol{C} \boldsymbol{X}_t^*(x_1, x_2), \end{aligned}$$

where

$$\begin{aligned} \mathrm{d}\, \boldsymbol{X}_t^*(x_1, x_2) &= \mathrm{d} \left(\frac{1}{x_2 - x_1} \boldsymbol{A}^{-1} \left(e^{\boldsymbol{A}x_2} - e^{\boldsymbol{A}x_1} \right) \right) \\ &= \frac{1}{x_2 - x_1} \boldsymbol{A}^{-1} \left(e^{\boldsymbol{A}x_2} - e^{\boldsymbol{A}x_1} \right) \boldsymbol{A} \boldsymbol{X}_t \, \mathrm{d}\, t + \underbrace{\frac{1}{x_2 - x_1} \boldsymbol{A}^{-1} \left(e^{\boldsymbol{A}x_2} - e^{\boldsymbol{A}x_1} \right) \boldsymbol{B}}_{=: \boldsymbol{B}^*(x_1, x_2)} \mathrm{d}\, \boldsymbol{L}_t \\ &= \boldsymbol{A} \frac{1}{x_2 - x_1} \boldsymbol{A}^{-1} \left(e^{\boldsymbol{A}x_2} - e^{\boldsymbol{A}x_1} \right) \boldsymbol{X}_t \, \mathrm{d}\, t + \boldsymbol{B}^*(x_1, x_2) \, \mathrm{d}\, \boldsymbol{L}_t \\ &= \boldsymbol{A} \boldsymbol{X}_t^*(x_1, x_2) \, \mathrm{d}\, t + \boldsymbol{B}^*(x_1, x_2) \, \mathrm{d}\, \boldsymbol{L}_t. \end{aligned}$$

This is only one representation for the futures. We could have also set $\boldsymbol{C}^*(x_1, x_2) := \boldsymbol{C} \frac{1}{x_2 - x_1} \boldsymbol{A}^{-1} \left(e^{\boldsymbol{A}x_2} - e^{\boldsymbol{A}x_1} \right)$ and $\boldsymbol{A}$ and $\boldsymbol{B}$ as before. The difference becomes clearer if we fix x_2 and x_1 and if we model the futures by a MCARMA process. Then, as we have seen, there exist multiple state space representations. But for example in the echelon representation $\boldsymbol{C}$ is a matrix consisting of zeros and ones and thus taking the model with $\boldsymbol{C}^*(x_1, x_2)$ would destroy this simple structure. Therefore the first representation with $\boldsymbol{B}^*$ is preferable for the echelon representation. On the other hand for the controller canonical form the matrix $\boldsymbol{B}$ has such a simple structure and the second representation with $\boldsymbol{C}^*$ should be chosen instead. Until now our process is defined under the measure $\mathcal{Q}$, but for the estimation of real world data we need the physical measure $\mathcal{P}$. Therefore we apply a Esscher transform to the future process, such that the tempered-copula-stable process changes to a copula-stable process. By Theorem 2.11 we know that this will introduce an additional drift in the copula-stable process, and hence translates by (4.5) to a drift in the future prices. This new drift will be captured in our empirical data analysis by a trend component in the deterministic part Λ^*.

5 Simulation Study of the Estimation Procedures

5.1 The quasi maximum likelihood estimation

In this section I present four analysis for a MCARMA process with different driving noises and different sample sizes. We will start with a multivariate Gaussian noise. I have used the echelon parametrization with Kronecker indices $(2,2)$ as given in Example 3.9. In order to make this model identifiable from the spectral density Assumption 9 of Theorem 3.11, Schlemm et al. [30] proposed to fix the value of the transfer function at the origin, i.e. $H_\vartheta(0) = P_\vartheta(0)^{-1}Q_\vartheta(0) = H_0 \in \mathbb{R}^{2\times 2}$. For this simulation study I have chosen $H_0 = -\boldsymbol{I}_2$ from which we can derive that the "intercepts" $-A_1$ and K_1 of the matrix polynomials $P(z)$ and $Q(z)$ are equal. This has also the side effect that our parameter space is reduced by four dimensions. The final state space model is given by

$$\boldsymbol{A} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ \vartheta_1 & \vartheta_2 & \vartheta_3 & \vartheta_4 \\ 0 & 0 & 0 & 1 \\ \vartheta_5 & \vartheta_6 & \vartheta_7 & \vartheta_8 \end{pmatrix}, \boldsymbol{B} = \begin{pmatrix} \vartheta_9 & \vartheta_{10} \\ \vartheta_1 + \vartheta_4\vartheta_{11} + \vartheta_2\vartheta_9 & \vartheta_3 + \vartheta_2\vartheta_{10} + \vartheta_4\vartheta_{12} \\ \vartheta_{11} & \vartheta_{12} \\ \vartheta_5 + \vartheta_8\vartheta_{11} + \vartheta_6\vartheta_9 & \vartheta_7 + \vartheta_6\vartheta_{10} + \vartheta_8\vartheta_{12} \end{pmatrix} \text{ and } \boldsymbol{C} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

By using an Euler approximation scheme we have simulated 200 different sample paths with the parameters $\vartheta_0 = (-0.1, -1.2, 0.0, 0.2, 1.5, 1.4, -0.75, -1.75, -2.7, 1.75, -3.5, 3.0, 4.5, 5.5, 7.0)$, where the last three parameters correspond to the covariance matrix of the Lévy noise. The Euler approximation is given for each $k = 1, ..., M/\Delta t_E$ by

$$\boldsymbol{X}_{k\Delta t_E} = \boldsymbol{X}_{(k-1)\Delta t_E} + \boldsymbol{A}_\vartheta \boldsymbol{X}_{(k-1)\Delta t_E} + \boldsymbol{B}_\vartheta \boldsymbol{L}_{(k-1)\Delta t_E}, \qquad \boldsymbol{Y}_{k\Delta t_E} = \boldsymbol{C}_\vartheta \boldsymbol{X}_{k\Delta t_E},$$

where Δt_E is the mesh size of the Euler scheme and were I have set $\boldsymbol{X}_0 := (0,0,0,0)'$. The mesh of the simulation time grid has been chosen to be $\Delta t_E = 0.0001$ and the total number of time steps is equal to 200000. However the observed output process is only every 100th simulated value and hence we observe the process at $M = 2000$ time points which are equally spaced with mesh size $\Delta t = 0.1$. To increase the speed of the optimization of the loglikelihood I have decided to chose the true parameter ϑ_0 as the initial point of the optimization since we will see that even for this starting point there are better parameters in terms of a lower log likelihood which are not even close to ϑ_0. Another advantage of this initial value is that, since it is well know that a QMLE can have many local minima, we do not end up in such a local minimum which has a higher log likelihood than the true value. Hence there is also no need for global optimization procedures. For the optimization of the log likelihood I have chosen a constraint interior-point routine, since this has performed best in terms of convergence speed and accuracy. In each iteration of the minimization routine, it is checked if the following constrains are fulfilled

1. The real parts of the eigenvalues of $\boldsymbol{A}$ are strictly negative
2. The imaginary parts of the eigenvalues of $\boldsymbol{A}$ are in $\left(-\frac{\pi}{\Delta t}, \frac{\pi}{\Delta t}\right)$
3. The state space system is minimal with McMillan degree equal to the raw size of $\boldsymbol{A}$.

Furthermore for the solution of the Riccati equation to determine Ω_ϑ, we used the MATLAB function "dare". The results of the QMLE can be found in Table 1. We observe that certain estimates, eg. ϑ_1 and ϑ_2, do not even give a reasonable estimate on average let alone their standard deviation. We should note here that these results are not in line with the results of the simulation study of Schlemm et al. [30], where the estimates for the same sample size, but with lower dimensional MCARMA process, were exact on average up to the second decimal place. To obtain a comparable result we had to increase the sample size by the factor of 100, i.e. 200000 observations of the MCARMA process. The detailed results can be found in Table 2.
In the next simulation I have chosen a copula-stable driving process and 2000 observations of the MCARMA process. Note that in this framework we do not have "true" values for the covariance matrix, since the driving process has no finite second moment. Therefore we interpret $\Sigma^{\boldsymbol{L}}$ as the empirical covariance matrix which can vary quite heavily. For the optimization procedure I calculated the empirical covariance of the copula-stable increments and used these as starting points. As we can see from Table 3 the estimates for $\vartheta_1, ..., \vartheta_8$ which corresponds to the matrix $\boldsymbol{A}$ are comparable to estimates in the Gaussian setup, whereas the estimates for the matrix $\boldsymbol{B}$ are totally useless. My explanation for this effect is, as mentioned earlier, that both $\boldsymbol{B}$ and $\Sigma^{\boldsymbol{L}}$ only appear in the calculation of the covariance of the state equation noise and that for a $T \in \mathbb{R}^{d\times d}$ such that $\boldsymbol{B}^* = \boldsymbol{B}T$ and $\Sigma^{*,\boldsymbol{L}} = T\Sigma^{\boldsymbol{L}}T'$ also results in an estimate with the same likelihood. For a small covariance matrix $\Sigma^{\boldsymbol{L}}$ this identification problem is less pronounced, but in this stable stimulation it can happen that $\Sigma^{\boldsymbol{L}}$ becomes very large and in these cases I also observed a large deviation in $\boldsymbol{B}$ and $\Sigma^{\boldsymbol{L}}$ from the true $\boldsymbol{B}$ and the empirical covariance of the simulated copula-stable increments. For comparison I present in Table 4 the QMLEs for almost the same setup, but now instead of estimating the empirical covariance matrix in the optimization routine, I fixed it to the empirical covariance of the simulated copula-stable increments. Furthermore, since this is the last simulation in this thesis, I also wanted to show that other state space representations work equally "well". Therefore the process I used in Table 4 was in controller canonical form and the parameters were chosen s.t. they are, up to two digits, the transformation of the previous parameters. I also checked the estimates in the echelon representation in this framework and they yielded a similar picture. After 30 simulations we can already see that the estimation for $\boldsymbol{B}$ is now comparable to the estimates of $\boldsymbol{A}$.

	ϑ_1	ϑ_2	ϑ_3	ϑ_4	ϑ_5	ϑ_6	ϑ_7	ϑ_8
True value	-0.1	-1.2	0	0.2	1.5	1.4	-0.75	-1.75
Sample mean	0.4183	-1.1890	-0.2667	-0.1810	1.5935	1.4056	-0.7972	-1.8180
Sample std. dev.	1.7276	0.1573	0.8639	1.2054	0.4538	0.0424	0.2272	0.3154
Sample min.	-3.2854	-1.8496	-4.9968	-6.5949	0.3994	1.2433	-1.7812	-3.1670
Sample max.	9.8062	-0.5915	1.6090	2.5115	3.5478	1.5204	-0.1989	-1.0110

	ϑ_9	ϑ_{10}	ϑ_{11}	ϑ_{12}	ϑ_{13}	ϑ_{14}	ϑ_{15}
True value	-2.7	1.75	-3.5	3	4.5	5.5	7
Sample mean	-3.0091	2.0369	-3.6686	3.2295	4.1989	5.0903	6.4693
Sample std. dev.	0.8831	0.7757	0.7900	0.7198	0.9521	1.2113	1.7059
Sample min.	-9.9626	0.6008	-7.6152	1.0472	1.0900	1.2594	1.5005
Sample max.	-1.4621	7.5962	-0.9336	6.4335	7.8707	8.7565	9.9833

Table 1: Statistics of the QMLEs for 200 sample paths of a MCARMA process with Gausian noise which is observed on the time horizon $[0, 200]$ with mesh size $\Delta t = 0.1$.

	ϑ_1	ϑ_2	ϑ_3	ϑ_4	ϑ_5	ϑ_6	ϑ_7	ϑ_8
True value	-0.1	-1.2	0	0.2	1.5	1.4	-0.75	-1.75
Sample Mean	-0.1286	-1.2120	0.0136	0.2248	1.5103	1.4006	-0.7551	-1.7564
Sample std. dev.	0.1844	0.0169	0.0919	0.1302	0.0588	0.0107	0.0295	0.0385
Sample min.	-0.6075	-1.2534	-0.1045	0.0574	1.4053	1.3718	-0.8197	-1.8303
Sample max.	0.1100	-1.1890	0.2516	0.5669	1.6382	1.4196	-0.7026	-1.6834

	ϑ_9	ϑ_{10}	ϑ_{11}	ϑ_{12}	ϑ_{13}	ϑ_{14}	ϑ_{15}
True value	-2.7	1.75	-3.5	3	4.5	5.5	7
Sample mean	-2.6942	1.7258	-3.5014	2.9801	4.4139	5.4322	6.9700
Sample std. dev.	0.0705	0.0748	0.0857	0.0840	0.1413	0.2074	0.3368
Sample min.	-2.8174	1.5628	-3.7174	2.8410	4.1547	4.9528	6.1496
Sample max.	-2.5092	1.8771	-3.3819	3.1866	4.8372	5.8432	7.6429

Table 2: Statistics of the QMLEs for 35 sample paths of a MCARMA process with Gausian noise which is observed on the time horizon $[0, 20000]$ with mesh size $\Delta t = 0.1$.

	ϑ_1	ϑ_2	ϑ_3	ϑ_4	ϑ_5	ϑ_6	ϑ_7	ϑ_8
True value	-0.1	-1.2	0	0.2	1.5	1.4	-0.75	-1.75
Sample mean	-0.1743	-1.2939	0.0309	0.2593	1.3968	1.2066	-0.7095	-1.6304
Sample std. dev.	0.6620	0.8123	0.3025	0.5422	1.2879	1.7204	0.5920	1.1406
Sample min.	-2.9172	-3.3818	-0.7792	-1.5693	-3.7934	-4.2428	-2.3880	-5.2628
Sample max.	1.5661	1.3220	1.2993	1.6212	4.9168	6.4422	1.6764	1.8990

	ϑ_9	ϑ_{10}	ϑ_{11}	ϑ_{12}	ϑ_{13}	ϑ_{14}	ϑ_{15}
True value	-2.7	1.75	-3.5	3	-	-	-
Sample mean	-1.3457	2.4707	0.1133	4.9611	48.8636	32.6357	189.1813
Sample std. dev.	4.7313	2.2010	8.7472	5.4433	176.1020	170.6384	541.5739
Sample min.	-16.4115	-2.1250	-8.8882	-3.5027	1.5017	-582.8710	0.1745
Sample max.	28.7879	13.3834	47.4530	33.3626	1318.6980	1052.7354	3239.1747

Table 3: Statistics of the QMLEs for 100 sample paths of a MCARMA process with $(1.7, 1.5)$-copula-stable noise which is observed on the time horizon $[0, 200]$ with mesh size $\Delta t = 0.1$.

	ϑ_1	ϑ_2	ϑ_3	ϑ_4	ϑ_5	ϑ_6	ϑ_7	ϑ_8
True value	1.75	-1.4	1.6	-1.25	1.5	-0.15	-0.05	1.45
Sample mean	1.7875	-1.4558	1.6779	-1.2881	1.5365	-0.2301	-0.1101	1.3890
Sample std. dev.	1.0961	1.3394	0.9077	1.1833	0.8291	0.9286	0.6327	0.6578
Sample min.	0.3333	-4.1759	0.3706	-3.5449	0.6278	-2.3364	-0.9694	-0.4123
Sample max.	3.9196	0.1523	3.4171	0.3223	3.3864	0.5930	1.5661	2.0700

	ϑ_9	ϑ_{10}	ϑ_{11}	ϑ_{12}
True value	-2.7	1.75	-3.5	3
Sample mean	-2.6982	1.5371	-3.5768	2.7140
Sample std. dev.	0.2204	0.5784	0.4549	0.8302
Sample min.	-3.4011	0.2856	-4.9732	1.1057
Sample max.	-2.2312	2.3386	-2.8047	4.3845

Table 4: Statistics of the QMLEs for 30 sample paths of a MCARMA process in controller canonical form with $(1.7, 1.5)$-copula-stable noise which is observed on the time horizon $[0, 200]$ with mesh size $\Delta t = 0.1$. The empirical covariance matrix parameters are fixed to the covariance of the simulated noise.

5.2 Recovery of the Lévy increments

I simulated various MCARMA processes in the canonical controller form with different MCARMA parameters and different Lévy noises. Afterwards I compared the estimated Lévy increments with the original ones from the simulation via their histograms. I could not even observe one process, where the estimated Lévy increments were out of proportion. Also for small sample sizes like 2000 observations the estimates were convincing, e.g. for two independent marginal stable processes the original stable parameters were within the 95% confidence interval of the estimated stable distribution which were calculated via the program "stable". Therefore the estimation procedure described in 3.4 works very well. An example for 20000 observations of a MCARMA process and $\Delta t = 0.1$ is given in Figure 5.1.

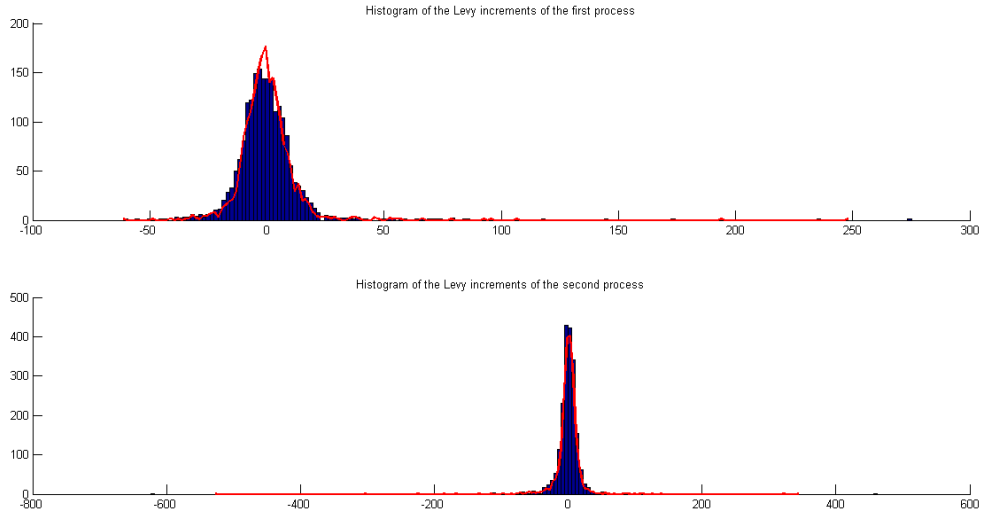


Figure 5.1: Histogram of a simulated (red) and the estimated (blue) Lévy increments of MCARMA process with $(1.7, 1.5)$-copula-stable noise with Clayton parameters $\delta = 1$ and $\eta = 0.5$. The marginal lévy measures are parametrized by $(c_1^+, c_1^-) = (1, 0.75)$ and $(c_2^+, c_2^-) = (0.75, 1)$. Furthermore I substracted the sample mean from the original Lévy increments.

6 Analysis of Electricity Spot Prices

Although the model of section 4 can be applied in general to all futures, we will use only the so called spot price of electricity data, which is also a future. In other markets the spot price refers to the price you have to pay to obtain the a certain product instantanously, whereas, due to transmission constrains and the limited storability of excess energy, the electricity spot market is an auction market, where offers and bids are placed for specific hours of the next day. At the European Energy Exchange (EEX) in Leipzig it is also possible to trade energy futures for the next weeks, months, quarters and years. The delivery period of those contracts start on the first day of the specific period, e.g. the first day of the month. The disadvantage arising from these types of contracts is that there is not a new contract every day. Assume that today is January 15th than we have a monthly future whose delivery period starts on February 1st, i.e. a time to maturity of 17 days, but on the next day we cannot observe the price of a future with a time to maturity of 17 days, but only one with 16 days. Therefore the next time we observe a similar future is in one month. As a result of this our data set would be quite small, hence we will work with at least daily futures which is the spot price. Because as we have seen in the prior simulation study the QMLE is only useful for a large data set, I have decided to use to electricity prices as the basis for the analysis, since these prices can also be observed on an hourly basis and on weekends, whereas the hourly spot prices will give rise to an additional problem.
Our data comprises hourly electricity prices from the German (EEX) and Scandinavian (Nordpool) market from January 1st, 2010 to December 15th, 2012 and daily electricity prices from the German and Italian (GME) market from January 2nd, 2008 to October 30th, 2012. Therefore we have 25920 hourly and 1765 daily observations for each corresponding market. The prices are displayed in Figure 6.2 and Figure 6.1.

First we start with the estimation of the seasonal component. It is common in the literature for electricity pricing to use a deterministic function consisting of a linear trend and a sum of trigonometric functions. On a first glance this type of seasonal function seems to be appropriate for the hourly data set, whereas our daily series starts in the middle of the financial crisis, which caused a sharp drop in the industrial production and thus also in the electricity price. Therefore a simple linear trend does not seem to be suited in this case. A different approach, which, as we will see later, works quite well for the daily spot prices, is an exponentially weighted moving average (EWMA) component in conjunction with an intercept and a sine/cosine function, which captures the regular yearly periodicity. Moreover we will also observe strong weekly seasonality for which we will use the standard tool in time series analysis, i.e. centered estimates for the mean of each weekday. To put it in mathematical terms we will model the seasonal component for daily observations of the spot price as

$$\Lambda(t) = \Lambda_1(t) + \Lambda_2(t) = c_1 + c_2 \sin(\frac{2\pi t}{365.25}) + c_3 \cos(\frac{2\pi t}{365.25}) + c_4 EWMA_t^{0.975} + s_t^w,\ c_1, ..., c_4 \in \mathbb{R}$$

where the exponentially weighted moving average is formally defined by $EWMA_t^a = (1 - a)f^*(t) + aEWMA_{t-1}^a$ for $a \in [0, 1]$. Furthermore the weekly seasonality for each workday

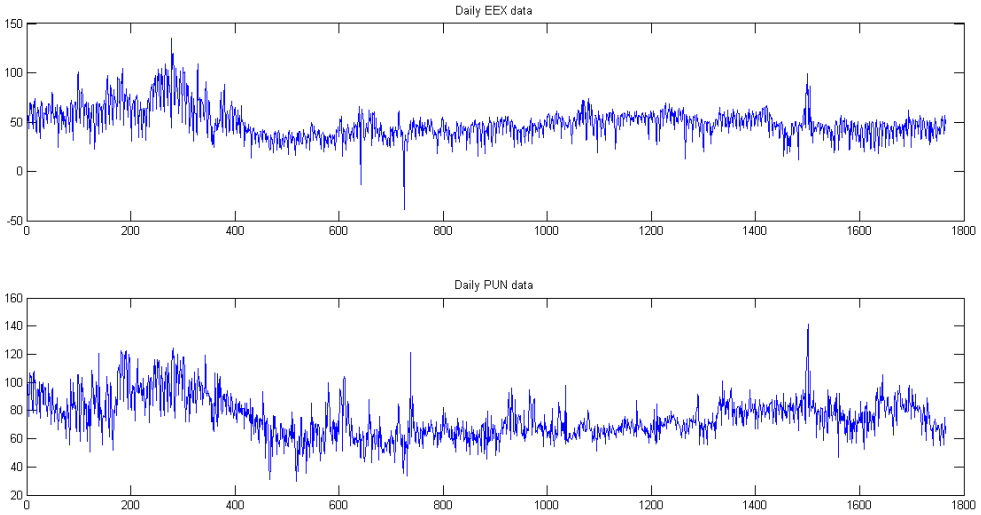


Figure 6.1: Daily spot prices.

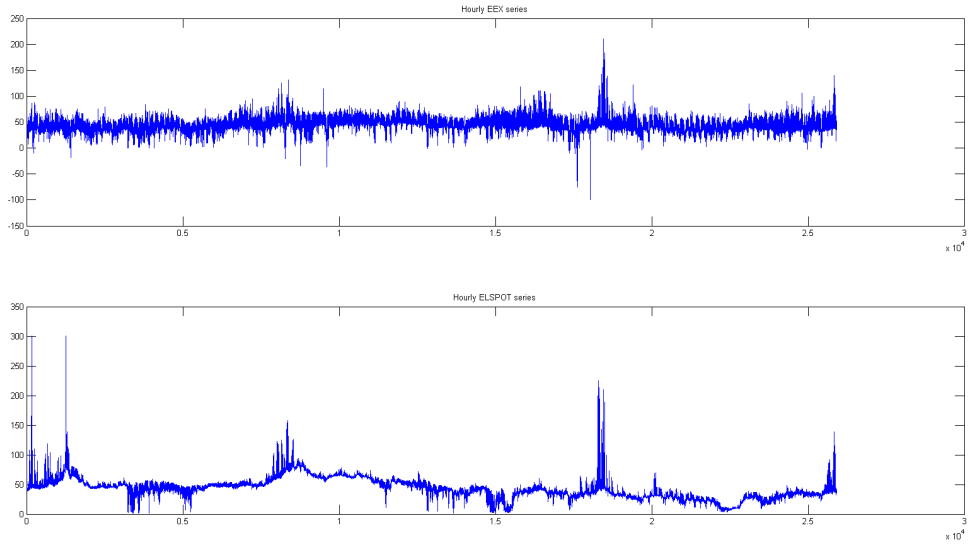


Figure 6.2: Hourly spot prices.

$r = 1, ..., 7$ is given by

$$s_r^w = s_r - \frac{1}{7}\sum_{j=1}^{7} s_j$$

and

$$s_r = \left\lfloor \frac{7}{M} \right\rfloor \sum_{j=1}^{\lfloor M/7 \rfloor} f^*\left(r + 7(j-1)\right).$$

where M denotes as usual the sample size and $f^*(t)$ is the observed spot price at time t with time to maturity of one day.
There a two main concerns using the seasonality function above. Firstly, due to the EWMA in the seasonality component, which depends on the current spot price, it is hard to project the seasonality into the future. Secondly, since energy can only be stored up to a limited degree, electricity prices tend to have spikes from time to time. This can be seen for example in the Scandinavian spot price series, where those spikes occur most of the time in the winter months. Those spikes will influence our estimates of the seasonality and thus we have to filter them out in order to obtain a robust result. A list of many filters can be found in [22]. Most of the filters concentrate on detecting all spikes, whereas our goal is to find the a robust seasonal function. Therefore we will use a recursive filter, whose stop criterion is based on the improvement of the mean squared error (MSE) of the estimation procedure. The basic algorithm is:

Algorithm 6.1 (Recursive filter)**.**

- Set $\alpha \in [0,1]$: Truncation level
- Set $\beta > 0$: Stopping condition
- Set $C = \beta + 1$
- Estimate Λ_1 from the observed series
- Save MSE and residuals
- WHILE $C > \beta$
  - Find the largest and lowest $\frac{\alpha}{2}\%$ of the residuals
  - Replace the corresponding values in the observed series by the median of their current month
  - Estimate Λ_1 from the new series
  - Save MSE and residuals
  - Calculate $C = 1 - \frac{MSE_{new}}{MSE_{old}}$
- END WHILE

In the analysis of the spot prices I extended the algorithm such that it also checks if after removing only the largest or smallest residuals the change in the MSE is at least $\frac{\beta}{2}$, otherwise it does not remove those residuals. I did this because for instance in the ELSPOT data most spikes are upward jumping and the basic algorithm would then also truncate the small downward jumps which do change the MSE by a large amount. The results for the estimation can be found in Figure 6.3 and Table 5.

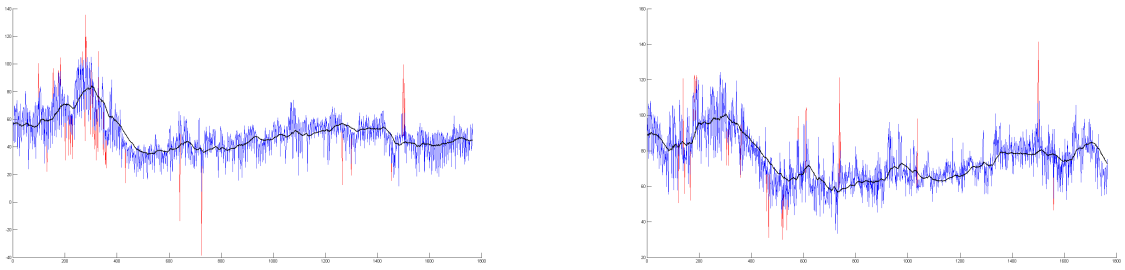


Figure 6.3: Daily spot prices (blue) without the removed spikes (red) via the recursive filter with $\alpha = 3\%$ and $\beta = 2.5\%$ and the estimated long term seasonal component Λ_1 (black). The upper figure corresponds to EEX and the lower to PUN data.

	c_1	c_2	c_3	c_4
EEX	-1.0771	-1.8401	-1.6840	1.0146
PUN	0.0064	-1.3678	-1.8744	0.9941

Table 5: Robust estimates for Λ_1. All estimates except the intercept c_1 of the PUN times series are significant.

If we look at the autocorrelation function of the residuals (Figure 6.4), we can clearly see a weekly seasonality for both time series. Therefore it is justified to include an additional weekly periodicity in the seasonality component. At first I also tried to deseasonalize the residuals via a trigonometric function, but the effect on the autocorrelation function was rather small. Therefore I decided to use the above described method for s^w. The estimates for s^w can be found in Table 6. Here we also see why the estimation with a sinodious function failed. We observe that the mean level of the EEX spot prices is almost constant in the middle of the week, whereas we see large deviations at the weekends and the sum of a single sine and cosine function is not able to model such fluctuations. We could still use the trigonometric approach, but then we have to include more sine/cosine functions with different periods and hence the space of estimated parameters would increase.

Finally the autocorrelation function as well as the Fourier amplitude did not show any further seasonalities.

Now we turn to the estimation of the seasonality function of the hourly spot prices. And

	s_1^w (Sat)	s_2^w (Sun)	s_3^w (Mon)	s_4^w (Tue)	s_5^w (Wed)	s_6^w (Thu)	s_7^w (Fri)
EEX	-6.0993	-11.3992	2.9337	4.2953	4.2248	4.0508	1.9940
PUN	-2.3837	-7.2740	1.6247	2.4435	2.7403	1.9400	0.9093

Table 6: Estimated centered daily medians per weekday.

to state the results right up front, I was not able to deseasonalize the time series. This is

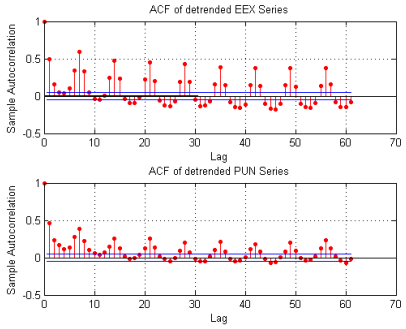


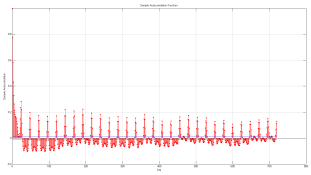
Figure 6.4: Autocorrelation functions of the daily spot prices without the estimated long term seasonal component Λ_1.

mainly due to the fact that the hourly spot prices are not time series, since the prices for the next 24 hours are calculated at the same time and thus no new information enters into the price from hour to hour. At first I used the same estimators as for the daily spot prices, with the exception that the weekly seasonality now comprises $24 \cdot 7 = 168$ estimates. But the autocorrelation function of the data sets show that I was not able to get rid off the intra-day seasonality. In the later approaches I tried to fit seasonality functions to the data, which consist of many trigonometric functions or dummy variables - for certain hours, days or months - or a mixture of both. Although I could reduce the seasonality with these functions, the intra-day seasonality was always present. Furthermore in contrast to the first method the autocorrelation function stays also for very large lags on a high level, which indicates an additional trend component. Because of this poor estimation results, it makes no sense to estimate the MCARMA parameters for this data set. Figure 6.5 shows the empirical autocorrelation functions for the hourly spot prices for the analogous method of the daily spot prices and for a seasonality function with many trigonometric functions.

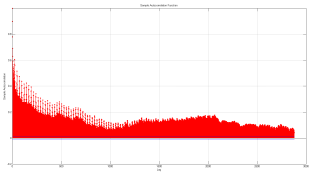
Now we start with the most delicate part, that is the estimation of the MCARMA parameters. For this purpose I propose a new method. We have seen in the simulation studies that the QMLE performs very bad in the infinite variance case when we also estimate the empirical covariance at the same time. Contrary the recovery procedure for the driving noise works very well. Therefore it seems natural to recover the Lévy noise for each ϑ at first and then to use the empirical covariance of the estimated Lévy noise in the QMLE. Since we have to recover the Lévy noise in each evaluation of the objective function the optimization algorithm is considerably slower. Moreover since we want to avoid the transformation from the left to the right matrix description, which is needed for the estimation of the Lévy noise, I used the controller canonical form in the QMLE procedure. Since I use a right matrix fraction description, I also take the state space form of our model where the time to maturity adjustment is located in the observation equation, that is the model with

$$\boldsymbol{C}^\star(x_1, x_2) = \boldsymbol{C} \frac{1}{x_2 - x_1} \boldsymbol{A}^{-1} \left(e^{\boldsymbol{A} x_2} - e^{\boldsymbol{A} x_1} \right).$$

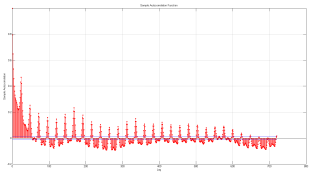
The daily spot price is calculated as the weighted average of the hourly spot prices which are



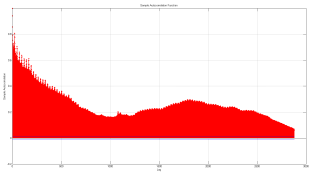
(a) EEX series with EWMA seasonality function



(b) EEX series with trigonometric seasonality function



(c) ELSPOT series with EWMA seasonality function



(d) ELSPOT series with trigonometric seasonality function

Figure 6.5: Estimated autocovariances for the deseasonalized hourly spot prices.

calculated at noon based on the results of the auction of this trading day. For example in the EEX Day Ahead market the auction takes place at 12:00 pm and thus the delivery period of the first future is 12 to 13 hours later, whereas for the last future the delivery period is 35 to 36 hours ahead.
For our data set I assumed that the times to maturity of delivery period are $x_1 = 1$ and $x_2 = 2$. These times to maturity of the delivery period are based on the assumption that the trading volumes of electricity for each delivery hour of the next day are symmetrical around the time of their price derivation.
Since x_1 and x_2 are fixed, I will treat the matrix $\boldsymbol{C}^*$ as the usual observation matrix in a MCARMA state space model. Afterwards we get an estimate for $\widehat{\boldsymbol{C}}$ from the estimated matrix $\widehat{\boldsymbol{C}}^*$ by

$$\widehat{\boldsymbol{C}} = \widehat{\boldsymbol{C}}^*(1,2)\left(e^{2\boldsymbol{A}} - e^{\boldsymbol{A}}\right)^{-1}\boldsymbol{A}.$$

Note that this inversion is always possible, since the eigenvalues of $\boldsymbol{A}$ have strictly negative real parts. To reduce confusion with all the state space representations I state the results of the estimation in terms of the AR and MA matrix polynomials P and Q.

$$\widehat{P}(z) = \boldsymbol{I}_2 z^2 + \begin{pmatrix} 1.3200 & -0.2090 \\ -0.3030 & 1.3900 \end{pmatrix} z + \begin{pmatrix} 0.0627 & 0.0047 \\ -0.0335 & 0.0814 \end{pmatrix},$$

and

$$\widehat{Q}(z) = \begin{pmatrix} -0.2980 & 0.0039 \\ 0.1340 & -0.4420 \end{pmatrix} z + \begin{pmatrix} -0.0627 & -0.0047 \\ 0.0335 & -0.0814 \end{pmatrix}$$

That the intercept matrices of $\widehat{P}$ and $\widehat{Q}$ are equal with different leading sign is not a coincidence. This is a consequence of the identifiability condition on the transfer function, where I

have chosen $H(0) = -\boldsymbol{I}_2$ once again. The empirical covariance function for the driving noise of the estimated MCARMA model is given by

$$\widehat{\Sigma}^{\widehat{\boldsymbol{L}}} = \begin{pmatrix} 997.74 & 386.75 \\ 386.75 & 780.07 \end{pmatrix}.$$

Furthermore histograms of the estimated Lévy increments with the estimated stable probability density functions are display in Figure 6.6, where the stable parameters were estimated by the program "stable". The parameters of the estimated marginal distributions can be found in Table 7.
In a final step I compared the empirical auto- and crosscorrelations with its theoretical counterparts for a finite variance driving Lévy noise, since theoretical auto- and crosscorrelations do not exist in the infinite variance framework. As we can see in Figure 6.7 our estimated MCARMA process approximates the empirical auto- and crosscorrelations quite well. It should be mentioned here that, although I have minimized the log likelihood with various global optimization routines, e.g. a simple patternsearch or the minimizer from the simulation study with 150 different initial points, I have never seen a similar result for a MCARMA process which was estimated simultaneously with the covariance of the Lévy noise.

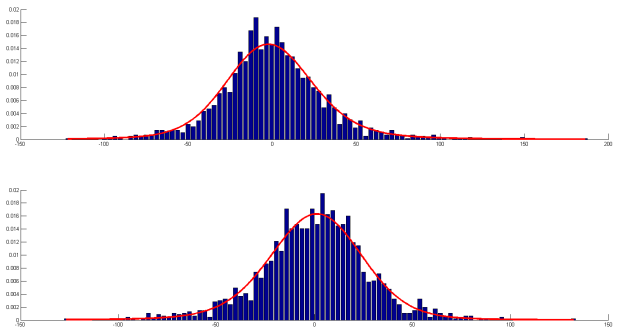
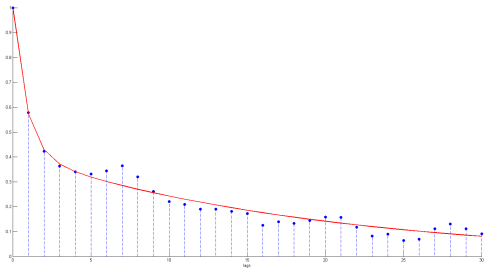


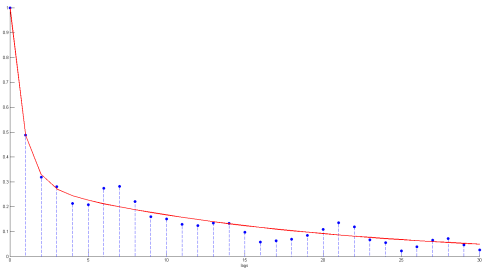
Figure 6.6: Histogram of the estimated Lévy increments of the MCARMA process for the daily spot prices. The red curve corresponds to the estimated pdf of the marginal stable distributions. The histogram on top/bottom corresponds to the first/second component of the driving noise.

	α	c^+	c^-	γ
$L^{(1)}$	1.6969	85.8835	54.0033	0.2596
$L^{(2)}$	1.7765	61.0394	98.1106	-0.2329

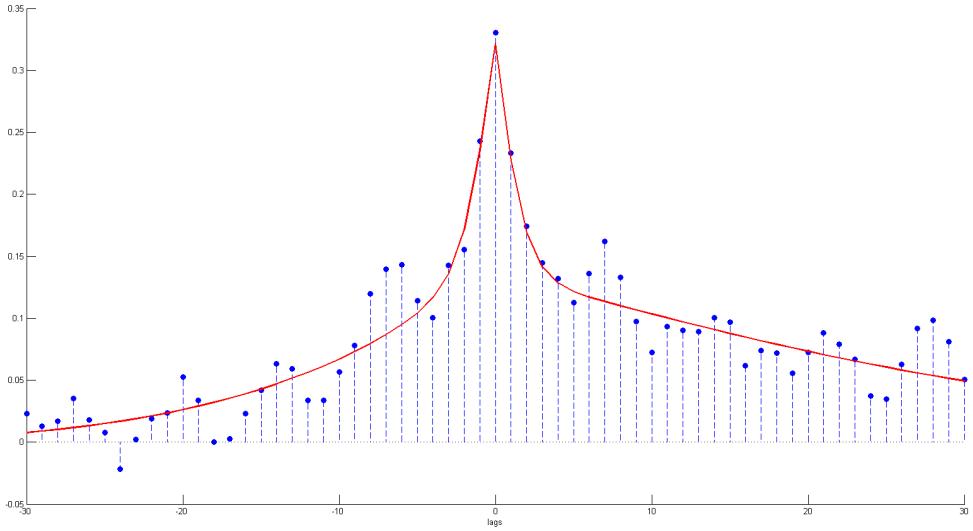
Table 7: Estimated marginal stable parameters of the driving noise. The parametrization is in line with Example 2.3.



(a) ACF of deseasonalized EEX series



(b) PUN of deseasonalized EEX series



(c) Crosscorrelations deseasonalized EEX and PUN series

Figure 6.7: Empirical (blue) and theoretical (assuming a finite variance driving noise; red) auto- and crosscorrelations for daily electricity spot prices.

7 Conclusion

In this thesis I investigated the multivariate Heath-Jarrow-Morton-Musiela model for energy futures. I have shown that the model can be represented as a sum of a multivariate CARMA process with copula-stable background driving noise and a seasonality function under the physical measure. The most time consuming part was the estimation of the MCARMA parameters, since the accuracy of the quasi maximum likelihood estimator is very low. I found out that this low precision is induced by estimating the covariance matrix of the driving Lévy process and the MCARMA parameters simultaneously. A way out of this problem is to use the estimated covariance matrix of the background noise. To this end we have to recover the noise in each iteration of the optimization routine of the QML. Although this procedure is very time consuming the result outperforms the simultaneous estimation method. Unfortunately this idea came to me in the final phase, otherwise I could have made a simulation study to test the superiority of the procedure. The next setback was that I was not able to estimate the Lévy copula parameters.

Acknowledgements

I would like to express my gratitude to Prof. Claudia Klüppelberg and Prof. Fred Espen Benth for their support and the pleasant stay in Oslo.

References

[1] Basilio, J.C.; Kouvaritakis, B. (1997) *An Algorithm for Coprime Matrix Fraction Description using Sylvester Matrices*, Linear Algebra and its applications Vol. 266, pp.107-125

[2] Benth, Fred E.; Koekebakker, Steen (2008) *Stochastic modeling of financial electricity contracts*, Energy Economics, Vol. 30, pp.1116-1157

[3] Benth, Fred E.; Klüppelberg, Claudia; Müller, Gernot; Vos, Linda (2012) *Futures pricing in electricity markets based on stable CARMA spot models*

[4] Bernstein, Dennis S. (2008) *Matrix mathematics: theory, facts, and formulas*, Princeton University Press

[5] Björk, Tomas; Gombani, Andrea (1999) *Minimal realizations of interest rate models*, Finance and Stochastics, Vol. 3, pp.413-432

[6] Brace, Alan; Musiela, Marek (1994) *A multifactor gauss markov implementation of Heath, Jarrow, and Morton*, Mathematical Finance, Vol. 4, pp.259-283

[7] Brockwell, Peter J.; Davis, Richard A. (1987) *Time Series: Theory and Methods*, Springer

[8] Brockwell, Peter J. (2001) *Lévy-Driven CARMA Processes*, Ann. Inst. Statist. Math., Vol. 53, No. 1, pp.113-124

[9] Brockwell, Peter J.; Lindner, Alexander M. (2009) *Existence and Uniqueness of Stationary Lévy-driven CARMA Processes*, Stoch. Proc. Appl. 119, pp.2660-2681

[10] Brockwell, Peter J.; Schlemm, Eckhard (2011) *Parametric estimation of the driving Lévy process of multivariate CARMA processes from discrete observations*

[11] Cohen, Serge; Rosinski, Jan (2007) *Gaussian approximation of multivariate Lévy processes with applications to simulation of tempered stable processes*, Bernoulli Vol.13, pp. 195-210

[12] Cont, Rama; Tankov, Peter (2004) *Financial Modelling with Jump Processes*, Chapman & Hall

[13] Davis, Richard A. (1996) *Gauss-Newton and M-estimation for ARMA processes with infinite variance*, Stoch. Proc. Appl. 63, pp. 75-95

[14] Didier, Sornette; Ide, Kayo (2001) *The Kalman-Lévy filter*, Physica D 151, pp. 142-174

[15] Esmaeili, Habib; Klüppelberg, Claudia (2012) *Two-step estimation of a multivariate Lévy process*, TU München

[16] Esmaeili, Habib; Klüppelberg, Claudia (2010) *Parametric estimation of a bivariate stable Lévy process*, TU München

[17] Fasen, Vicky; Fuchs, Florian (2012) *Spectral Estimation of High-Frequency Sampled CARMA Processes*

[18] Fuchs, Forian; Stelzer, Robert (2011) *Spectral Representation of Multivariate Regularly Varying Lévy and CARMA Processes*, to appear in Journal of Theoretical Probability

[19] Garcia, Isabel; Klüppelberg, Claudia; Müller, Gernot (2010) *Estimation of stable CARMA models with an application to electricity spot prices*, Ruhr-Universität Bochum

[20] Guidorzi, Roberto P. (1989) *Equivalence, Invariance and Dynamical System Canonical Modelling*, Kybernetika Vol. 25, No. 4, pp.233-257

[21] Hille, Einar; Phillips, Ralph S. (2000) *Functional Analysis and Semi-Groups*, AMS, Vol. 31

[22] Janczura, Joanna; Trück, Stefan; Weron, Rafal; Wolff, Rodney (2012) *Identifying spikes and seasonal components in electricity spot price data: A guide to robust modeling*, MPRA Paper No. 39277

[23] Kallsen, Jan; Tankov, Peter (2006) *Characterization of dependence of multidimensional Lévy processes using Lévy copulas*, Journal of Multivariate Analysis, Vol. 97, pp.1551-1572

[24] Marquardt, T.; Stelzer, R. (2007) *Multivariate CARMA processes*, Stoch. Proc. Appl. 117, pp. 96-120

[25] Protter, Philip E. (2004) *Stochastic Integration and Differential Equations*, Springer Verlag

[26] Rosiński, Jan (2007) *Tempering stable processes*, Stoch. Proc. Appl. 117, pp.677-707

[27] Rugh, Wilson J., (1996)*Linear System Theory*, Prentice Hall

[28] Sato, Ken-Iti (1999) *Lévy Processes and Infinitely Divisible Distributions*, Cambridge University Press

[29] Samorodnitsky, Gennady; Taqqu, Murad S. (2000) *Stable Non-Gaussian Random Processes*, Chapman&Hall

[30] Schlemm, Eckhard; Stelzer, Robert (2011) *Quasi maximum likelihood estimation for strongly mixing state space models and multivatiate Lévy-driven CARMA processes*, TU München

[31] Schlemm, Eckhard; Stelzer, Robert (2012) *Multivariate CARMA Processes, Continuous-Time State Space Models and Complete Regularity of the Innovations of the Sampled Processes*, Bernoulli 18.1, pp. 46-63

[32] Tankov, Peter (2004) *Lévy processes in Finance: Inverse Problems and Dependance Modelling*, PhD thesis