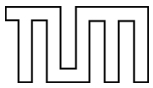


Technische Universität München

Department of Mathematics

# Modelling Temperature and Pricing Weather Derivatives

Master Thesis

by

Siting Huang

| | |
|---|---|
| Supervisor: | Prof. Dr. Claudia Klüppelberg |
| Advisor: | Dr. Jianing Zhang |
| Submission Date: | 2. January 2014 |

I hereby declare that this thesis is my own work and that no other sources have been used except those clearly indicated and referenced.

Munich, 2. January 2014

# Acknowledgments

First I want to express my gratitude to Prof. Dr. Claudia Klüppelberg for offering me an interesting topic and giving me the chance to write my master thesis at the Chair of Mathematical Statistics at the Technische Universität München. I also want to thank her for her helpful advice and encouragements.

Special thanks to my supervisor Dr. Jianing Zhang for his guidance and the support with valuable ideas and instructions. I am grateful to his patience and help with all questions during the work for my master thesis.

Furthermore I want to thank my friends for their company and encouragements in my master study in Germany.

Last but not least, I would like to thank my family for their love and support all the time in my life.

**Abstract**

The Chicago Mercantile Exchange offers futures and options based on daily temperatures. There is an increasing interest in developing a pricing model for these temperature derivatives by modelling the underlying temperature dynamics. In this work, we model the daily average temperature using a Lévy driven continuous-time autoregressive process with seasonal volatility. The aim is to generalize the Brownian motion as the driving noise process to Lévy processes. The family of generalized hyperbolic distributions is considered in order to capture the heavy tails in the temperature data. Based on the proposed model, we then present explicit pricing formulas for several temperature derivatives. Fourier technique is applied for numerical estimation for some derivatives as well. In empirical analysis, we first examine the temperature data of 6 US cities and 2 European cities. Further we implement the pricing framework for futures and discuss the behavior of the futures prices.

# Contents

# List of Tables

# List of Figures

# Chapter 1

# Introduction

## 1.1 Market for Temperature Derivatives

The first weather transaction was executed in 1997 as a weather option embedded in a power contract. Later in 1999, the Chicago Mercantile Exchange (CME) launched the first exchange-traded weather derivatives contracts. [1] Since then, the weather market has grown tremendously.

The interest on weather derivatives is aroused by the significant influence of weather. Business in Energy, agriculture, transportation, retail sales, etc. are easily affected by weather. Undesired weather events might lead to great loss in revenues. For example, cost of heating is increased in cold days, farming will face loss in a flood, and companies selling skiing instruments make less profit in warm seasons. Weather derivatives serve as an important tool to manage weather related risks. On the other hand, they also provide opportunities to speculate for possible profit on weather variations. Hence they attract not only companies facing weather related risk but also other market participants.

Currently the CME offers products related to temperatures, snowfall, frost, rainfall and hurricanes. Among all types of weather instruments, temperature based contracts are the most liquid ones. Temperature derivatives are based on four types of temperature indices: the Heating Degree Days (HDD), the Cooling Degree Days (CDD), the Cumulative Average Temperature (CAT) and the Pacific Rim Index (PAC). These temperature indices are location specific, and in total cover 24 cities in the United States, 11 in Europe, 6 in Canada, 3 in Australia and 3 in Japan. Table 1.1 lists the products available on CME currently for each location.

Here we introduce four types of temperature indices in the temperature market. Most of them are based on the daily temperature. Let $T(t)$ be the daily average temperature

[1] `http://www.cmegroup.com/trading/weather/`

Table 1.1: Weather Products on CME

| U.S. | Canada | Europe | Asia-Pacific | Australia |
|---|---|---|---|---|
| CDD Monthly | CAT Monthly | CAT Monthly | Asia-Pacific Monthly | CDD Monthly |
| CDD Seasonal | CAT Seasonal | CAT Seasonal | Asia-Pacific Seasonal | CDD Seasonal |
| HDD Monthly | CDD Monthly | HDD Monthly | | HDD Monthly |
| HDD Seasonal | CDD Seasonal | HDD Seasonal | | HDD Seasonal |
| Weekly Weather | HDD Monthly | | | |
| | HDD Seasonal | | | |

(DAT) on date $t$ and defined as

$$T(t) = \frac{T_{max}(t) + T_{min}(t)}{2} \tag{1.1}$$

where $T_{max}(t)$, $T_{min}(t)$ are the maximum and minimum temperature measured on date $t$, respectively.

The CAT index simply measures the cumulative DATs over the measurement period $[\tau_1, \tau_2]$. It is defined as

$$CAT(\tau_1, \tau_2) = \sum_{t=\tau_1}^{\tau_2} T(t). \tag{1.2}$$

The measurement period covers all year around.

HDD and CDD are cumulative difference between DAT and a reference temperature, usually $18°C$ or $65°F$. The HDD index measures how many degrees the temperature is below a certain baseline, and thus assesses the demand for heating. The CDD index, on the contrary, measures how many degrees the temperature is above a baseline, and assesses the demand for cooling. The measurement period for HDD Monthly and HDD Seasonal is from October to April, while the measurement period for CDD Monthly and CDD Seasonal is from April to October. The HDD over a measurement period $[\tau_1, \tau_2]$ is defined as

$$HDD(\tau_1, \tau_2) = \sum_{t=\tau_1}^{\tau_2} \max(c - T(t), 0). \tag{1.3}$$

Similarly, the CDD over $[\tau_1, \tau_2]$ is defined as

$$CDD(\tau_1, \tau_2) = \sum_{t=\tau_1}^{\tau_2} \max(T(t) - c, 0). \tag{1.4}$$

The Pacific Rim Index measures the cumulative total of 24-hour average temperatures

for Japanese cities:

$$PAC(\tau_1, \tau_2) = \sum_{t=\tau_1}^{\tau_2} \tilde{T}(t), \tag{1.5}$$

where $\tilde{T}(t) = \frac{1}{24}\sum_{i=1}^{24} T_i(t)$ with $T_i(t)$ being the temperature of the $i$th hour on the day $t$.

A *futures contract*, or *futures*, is a standardized contract between two parties to buy or sell a specified asset for a price agreed today. The agreed price is named as the *futures price*. The CME offers futures written on temperature indices. More explicitly, a temperature futures is an agreement to exchange a fixed amount of money for a variable referenced index with delivery and payment occurring at a specified future date. The referenced index is some temperature index, such as the HDD, CDD or CAT index, with a specified measurement period. The futures contracts are cash settled. The buyer of the contract receives 20 times the index value at the end of the measurement period in return for the futures price. The currency is Euro for European futures (British pounds for London) and US dollars for US futures. Trading of the futures are available up to the beginning of the measurement period.

The options offered in the CME are *plain vanilla European call and put options* written on the temperature futures. In this thesis, we simply refer a plain vanilla European call (or put) option as a call (or put) option for short. A *call option* (*put option*, respectively) is a contract that gives the holder the right, but not the obligation, to buy (sell, respectively) an underlying asset or instrument for a fixed price, known as the *strike price*, at a certain date in the future, known as the *expiration date* or *maturity*. Options written on futures are called *futures options*. The strike price of the option written on temperature futures is the specified futures price at which the underlying futures is traded. The options are cash settled. The owner of an option receives the difference between the futures price and the strike if the option is exercised. For example, suppose the futures price at exercise time $\tau < \tau_1$ is $F(\tau, \tau_1, \tau_2)$ for futures with measurement period $[\tau_1, \tau_2]$, then the owner of a call option receive the payoff

$$\max(F(\tau, \tau_1, \tau_2) - K, 0),$$

where $K$ is the strike price. In the market, we have options with different strike prices and exercise times for futures on different indices at all times, see Benth et al. [2007].

## 1.2 Pricing Temperature Derivatives via Daily Modelling

The daily modelling approach involves statistical modelling of the temperature evolution. One tries to model the dynamics of the temperature and then derive the dynamics of the indices and price the associating temperature derivatives under some theory of asset pricing. Instead of modelling the underlying temperature, one can model the dynamics of an index, named as the index modelling approach. Other common approaches are, for example, historical burn analysis and actuarial method.

The daily modelling approach has the following potential advantages, as argued in Jewson and Brix [2005]. First of all, the daily modelling preserves more information of the historical data. The daily modelling makes a more complete use of temperature data, while other approaches account for the temperature indices which include refinement of temperature data and lead to information lost. Secondly, daily modelling provides more accurate representation of all indices and their distributions. By daily modeling, the distribution of an index is derived from the distribution of temperature. More flexible and accurate distribution is then allowed compared to direct modelling of the index. Another advantage is that only one temperature model is required for all contracts on one location. This ensures necessary consistency in pricing different contracts on the same location, which will further suit the purpose of hedging. Compared to daily modelling, index modelling where different models for different indices on the same location are usually needed. Therefore, it is also more convenience in the sense of implementation. From the point of implementation, daily modelling also allows to incorporate meteorological forecasts, which can further increase the accuracy of forecasting and hence of pricing.

The disadvantage of daily modelling lies in the increased complexity of the model. Since temperature is a complicated system in its nature, it is not straightforward to develop a model fully capturing its dynamics. It is expected that the more sophisticated models we use for modelling temperature, the more complex it will be to fit and implement the models. Besides, daily modelling bears the risk from the model error. If the temperature model is not good enough, or a mistake occurs when implementing the model, we may end up with large error in pricing of the temperature derivatives. It is also worth mentioning that access to weather data can be costly and not always available.

An accurate yet practically manageable model for the temperature dynamics is required as the starting point for further steps of pricing. Different models have been proposed to model the dynamics of daily temperature. In general, they can be classified into two kinds: the discrete time models and the continuous time models. Since discrete values

of temperature are used in determining temperature indices, discrete time models are naturally adopted to study the daily observations. A general autoregressive moving average (ARMA) framework is often used in literatures considering discrete time formulation, see for example, Campbell and Diebold [2005], Roustant et al. [2004] and Caballero et al. [2002]. On the other hand, continuous time models attracts more and more attention in academic recently. From the point of modelling temperature, they are capable of describing irregularly spaced data and high-frequency data. From the point of pricing, they allows application of some modern finance theory where theoretical solutions in derivatives pricing can be obtained.

Many literatures attempt to model the temperature using mean-reverting Ornstein-Uhlenbeck processes, where Brownian motion is commonly considered as the driving noise process, see for example Choyce [2000] and Alaton et al. [2002]. However, Brownian motion is found out to be insufficient to model the evolution of the temperature. Brody et al. [2002] suggests using a fractional Brownian motion as the driving noise to incorporate long memory effect. The family of generalized hyperbolic Lévy processes are suggested in order to capture the semi-heavy tails and skewness of residuals in the temperature. For example, Bellini [2005] suggests a hyperbolic Lévy process; Alexandridis and Zapranis [2013] examines generalized hyperbolic distribution, the hyperbolic, the normal inverse Gaussian and stable distribution for the residuals.

Benth et al. [2007] generalizes the above Ornstein Uhlenbeck process to higher order continuous-time autoregressive (CAR) process with lag $p$ and seasonal variance to model temperature dynamics, where Brownian motion is the driving noise. Compared to Ornstein-Uhlenbeck process, which is a CAR process with lag 1, the CAR process with lag $p$ allows a more flexible choice for the autocorrelation structure in the daily temperatures. A generalization of this model is to generalize the Brownian motion to Lévy processes. For example, Swishchuk and Cui [2013] applies the Lévy driven CAR process and uses generalized hyperbolic distributions for the residuals to Canadian temperature data. The pricing formula for CAT futures and numerical pricing formula for CDD and HDD futures are also discussed in Swishchuk and Cui [2013].

## 1.3 Motivation and Outline

As introduced before, the CAR process provides an useful tool in the context of temperature derivatives. In this thesis, we explore the framework of modeling and pricing weather derivatives under a Lévy driven CAR process.

For the part of modelling the temperature, we use a CAR process with lag $p$ and sea-

sonal variance. Some properties of such processes are first studied. Typically, we introduce the class of Lévy processes with marginals following the family of generalized hyperbolic distributions, which is capable to describe the heavy tails observed in the data.

Then we proceed on pricing the weather derivatives under our temperature model. Since we consider contracts written on temperature indices which are in fact not traded in the market, we deal with an incomplete market model. The arbitrage theory for asset pricing is first reviewed. Under the no-arbitrage assumption, the derivatives prices are some expected values under the equivalent martingale measure. The technique of applying Esscher transform to construct risk-neutral measures is applied. We further introduce explicit pricing formulas for CAT, HDD and CDD futures, as well as CAT options under the settings of Brownian motion. For Lévy driven CAR process, explicit formula can be derived for only CAT futures price. For HDD futures, CDD futures and CAT options, the method of Fourier pricing is explored.

Further we implement the framework of pricing under the Lévy driven CAR model. First we investigate the CAR model on temperature data for 6 US cities and 2 European cities. Then we estimate the CAT futures prices and HDD prices for Stockholm following the derived formulas and Fourier approach.

The thesis is structured as follows. After this introductory chapter, the necessary background knowledge for both modelling and pricing is introduced in chapter 2. More exactly, basics of the Lévy processes, arbitrage theory, Esscher transform and time series analysis are included here. Chapter 3 presents the definition and some properties of the CAR processes, and specifies the CAR model for temperature. In chapter 4, issues of pricing temperature futures and options under the Lévy driven CAR model are discussed. The data analysis for temperature is shown in chapter 5. Empirical results of pricing are presented in chapter 6. Chapter 7 gives a summary of the thesis. All computations in this thesis are carried out in R version 3.0.2.

# Chapter 2

# Preliminaries

## 2.1 Lévy Process

Lévy processes focus on jump behaviors and they allow more realistic and flexible models to describe the evolution of temperature. Later we will assume Lévy processes as the driving noise in our model for temperature dynamics. The class of generalized hyperbolic Lévy processes are typically of interest since they provide distributions that are more close to empirical statistics of temperature data.

For the purpose of introduction, we first give the definition of a Lévy process and introduce its distribution. Definitions and some properties of the class of generalized hyperbolic Lévy processes are followed. Then we present some results on the stochastic integration based on Lévy process, which is needed for further discussion.

### 2.1.1 Lévy Process and Infinite Divisibility

Assume that a filtered probability space $(\Omega, \mathcal{F}, P)$ with filtration $\mathbb{F} = (\mathcal{F}_t)_{0 \le t \le \infty}$, denoted as $(\Omega, \mathcal{F}, \mathbb{F}, P)$ is given. Moreover, we always assume that it satisfies the *usual hypotheses*:

(i) $\mathcal{F}_0$ contains all the $P$-null sets of $\mathcal{F}$;

(ii) $\mathcal{F}_t = \cap_{u>t} \mathcal{F}_u$ for all $0 \le t < \infty$, i.e. the filtration $\mathbb{F}$ is right continuous.

A *stochastic process* $(X(t), t \in \mathcal{T})$ is a family of random variables defined on a probability space $(\Omega, \mathcal{F}, P)$. The index set $\mathcal{T}$ is called the *time domain*. Here we consider $\mathcal{T} = [0, \infty)$ or $\mathcal{T} = \mathbb{R}$ for continuous-time stochastic processes, and $\mathcal{T} = \mathbb{Z}$ or $\mathcal{T} = \mathbb{Z}^+$ for discrete-time ones. $\mathbb{R}$ denotes the set of real numbers, $\mathbb{R}^+$ the set of positive real numbers, $\mathbb{Z}$ the set of integers, $\mathbb{Z}^+$ the set of positive integers, $\mathbb{C}$ the complex plane.

The following definitions and results can be found in any standard text book on Lévy process. Here we mainly refer to Applebaum [2009].

**Definition 2.1** (Lévy process)**.** *A stochastic process $L = (L(t), t \geq 0)$ is a Lévy Process if*

*(i)* $L(0) = 0$ *almost surely;*

*(ii)* *L has independent increments, i.e. for each* $n \in \mathbb{N}$ *and each* $0 \leq t_1 \leq \cdots \leq t_{n+1} < \infty$ *the random variables* $(X(t_{j+1} - X(t_j)), 1 \leq j \leq n)$ *are independent;*

*(iii)* *L has stationary increments, i.e.* $X(t) - X(s) \stackrel{d}{=} X(t-s)$ *for all* $s \leq t$*;*

*(iv)* *L is stochastically continuous, i.e.* $\lim_{t \to s} P(|L(t) - L(s)| > a) = 0$*, for all* $a > 0$*, for all* $s \geq 0$

The Brownian motion and the Poisson process are two special cases of Lévy processes.

**Example 1** (Brownian motion)**.** *A (standard) Brownian motion is a Lévy process $B = (B(t), t \geq 0)$ for which*

*(i)* $B(t) \sim N(0, t)$ *for each* $t \geq 0$*,*

*(ii)* *B has continuous sample paths.*

In this thesis, we always use $B$ to denote a Brownian motion and $L$ to denote a Lévy process.

**Example 2** (The Poisson process)**.** *The Poisson process $N = (N(t), t \geq 0)$ of intensity $\lambda > 0$ is a Lévy process taking values in $\{0, 1, 2, \dots\}$ where $N(t)$ follows a Poisson distribution with intensity $\lambda t$, so that*

$$P(N(t) = n) = \frac{(\lambda t)^n}{n!} e^{-\lambda t}$$

*for each* $n = 0, 1, 2, \dots$*. The compensated Poisson process $\tilde{N} = (\tilde{N}(t), t \geq 0)$ is defined by $\tilde{N}(t) = N(t) - \lambda t$ for $t \geq 0$.*

The distributions of Lévy processes are are related to infinitely divisible distributions. We first give the definitions of infinitely divisible random variables and distributions.

**Definition 2.2** (Infinitely divisible)**.** *Let $X$ be a random variable taking values $\mathbb{R}$. We say $X$ is infinitely divisible if, for all $n \in \mathbb{N}$, there exits i.i.d random variables $Y_1^{(n)}, \dots, Y_n^{(n)}$ such that*

$$X = Y_1^{(n)} + \cdots + Y_n^{(n)}.$$

*A distribution $\mu(dz)$ is said to be infinitely divisible if there exits a distribution $\mu^{(n)}$ for each $n = 1, 2, \ldots$ such that*

$$\mu = \mu^{(n)} * \mu^{(n)} * \cdots * \mu^{(n)}, \tag{2.1}$$

*where $*$ is a convolution operator.*

Given a random variable $X$, let $\phi_X(u) = \mathbb{E}(e^{iuX})$ denote the characteristic function of $X$ where $u \in \mathbb{R}$ and $i = \sqrt{-1}$. The following Lévy-Khintchine Formula gives a characterisation of infinitely divisible random variables through their characteristic functions. First we need to introduce the Lévy measure. A *Lévy measure*, denoted by $\ell(dz)$, is a $\sigma$-finite measure on Borel sets of $\mathbb{R}\backslash\{0\} := \{x \in \mathbb{R}, x \neq 0\}$ satisfying the integrability condition

$$\int_{\mathbb{R}\backslash\{0\}} 1 \wedge z^2 \ell(dz) < \infty$$

Here $a \wedge b$ denotes the minimum of two numbers $a$ and $b$. The Lévy-Khintchine Formula follows

**Theorem 2.3** (Lévy-Khintchine)**.** *A random variable $X$ is a infinitely divisible is there exists $\nu \in \mathbb{R}$, $\sigma > 0$ and a Lévy measure $\ell$ such that, for all $u \in \mathbb{R}$,*

$$\phi_X(u) = \exp\left\{iu\nu - \frac{1}{2}\sigma^2 u^2 + \int_{\mathbb{R}} \left(e^{iuz} - 1 - iuz1_{|z|<1}(z)\right) \ell(dz)\right\}. \tag{2.2}$$

*Conversely, any mapping of the form (2.2) is the characteristic function of an infinitely divisible random variable on $\mathbb{R}$.*

**Remark 1.** *The characteristic function (2.2) can be written as $\phi_X(u) = e^{\eta(u)}$, where $\eta : \mathbb{R} \mapsto \mathbb{C}$,*

$$\eta(u) = iu\nu - \frac{1}{2}\sigma^2 u^2 + \int_{\mathbb{R}} \{e^{iuz} - 1 - iuz1_{|z|<1}(z)\}\ell(dz),$$

*is called the Lévy symbol. The triplet of parameters $(\nu, \sigma, \ell)$ is called the generating triplet.*

By the definition of a Lévy process $(L(t), t \geq 0)$,

$$L(t) = L(\frac{t}{n}) + (L(\frac{2t}{n}) - L(\frac{t}{n})) + \cdots + (L(t) - L(\frac{(n-1)t}{n})),$$

for each $n = 1, 2, \ldots$, so that $L(t)$ is the sum of $n$ independent, identically distributed (i.i.d) random variables, whose distribution is that of $L(\frac{t}{n})$. Therefore, $L(t)$ is infinitely divisible for each $t \geq 0$. By Lévy-Khintchine Formula, we have $\phi_{L(t)}(u) = e^{\eta(t,u)}$ for each

$t \geq 0$, $u \in \mathbb{R}$, where each $\eta(t, \cdot)$ is a Lévy symbol. We see in the following theorem that $\eta(t, u) = t\eta(1, u)$ for each $t \geq 0$, $u \in \mathbb{R}$.

**Theorem 2.4.** *If $L$ is a Lévy process, then the characteristic function of $L(t)$*

$$\phi_{L(t)}(u) = e^{t\eta(u)},$$

*for each $u \in \mathbb{R}$, $t \geq 0$, where $\eta(u)$ is the Lévy symbol of $L(1)$.*

Therefore, we define the Lévy symbol and the generating triplet of a Lévy process $(L(t), t \geq 0)$ to be those of $L(1)$. If $L(t)$ is a Lévy process, the distribution of $L(1)$ is infinitely divisible. Conversely, as stated in Applebaum [2009], page 62, Corollary 1.4.6,

**Theorem 2.5.** *Given an infinitely divisible probability measure $\mu$ on $\mathbb{R}$ with Lévy symbol $\eta$, there exists a Lévy process $\{L(t)\}$ such that $\mu$ is the law of $L(1)$.*

From the following Lévy-Itô decomposition we see that a Lévy process can be decomposed into a Brownian motion with drift (the continuous path), a Poisson integral (the large jump) and a compensated Poisson integral (the small jumps), cf. Applebaum [2009], page 108, Theorem 2.4.16,

**Theorem 2.6** (Lévy-Itô decomposition)**.** *If $L$ is a Lévy process, then there exists $\nu \in \mathbb{R}$, $\sigma > 0$ , a Brownian motion $B(t)$ and an independent Poisson random measure $N$ on $\mathbb{R}^+ \times (\mathbb{R}\backslash\{0\})$, such that for for each $t \geq 0$,*

$$L(t) = \nu t + \sigma B(t) + \int_{|z|<1} z\tilde{N}(t, dz) + \int_{|z|\geq 1} zN(t, dz),$$

*Here $N(dt, dx)$ is a Poisson random measure with compensator $dt \times \ell$, where $\ell$ is the Lévy measure and $dt$ denotes the Lebesgue measure, and $\tilde{N}(dt, dz) = N(dt, dz) - dt\ell(dz)$.*

The Lévy process $L(t)$ with above Lévy-Itô decomposition corresponds to a generating triplet $(\nu, \sigma, \ell)$. For example, if the generating triplet is $(\nu, \sigma, 0)$, then the corresponding Lévy process is $L(t) = \nu t + \sigma B(t)$, which is a Brownian motion with variance $\sigma^2$ and drift $\nu$.

In the case where $\sigma = 0$, i.e., the process has no continuous Brownian motion part, the process is a *purely discontinues Lévy process*. In next section, we introduce the class of generalized hyperbolic Lévy processes which belongs to this type.

### 2.1.2 Generalized Hyperbolic Lévy Processes

Now we introduce the family of generalized hyperbolic distributions and two of its subclasses: the hyperbolic distribution and the normal inverse Gaussian distribution. These distributions are infinitely divisible. As discussed before, they generate corresponding Lévy processes.

The generalized hyperbolic distribution is often used in modelling financial markets. It is a very flexible family of distributions which can model skewness and (semi-)heavy tails. This is a property that the normal distribution does not possess. Moreover, the density function, the characteristic function and the moment generating function of generalized hyperbolic distributions are explicitly known, which provides convenience to access pricing as we will see later.

The generalized hyperbolic distribution, introduced by Barndorff-Nielsen and Halgreen [1977], is a class of five-parameter distributions.

**Definition 2.7** (Generalized hyperbolic Lévy process)**.** *The generalized hyperbolic distributions, denoted by $GH(\lambda, \alpha, \beta, \mu, \delta)$, with parameters $\lambda \in \mathbb{R}$, $\alpha > 0$, $|\beta| < \alpha$, $\delta > 0$ and $\mu \in \mathbb{R}$, is a family of infinitely divisible distributions with density function given by*

$$f_{GH}(x; \lambda, \alpha, \beta, \mu, \delta) = \frac{\left(\sqrt{\alpha^2 - \beta^2}/\delta\right)^{\lambda} e^{\beta(x-\mu)}}{\sqrt{2\pi} K_{\lambda}\left(\delta\sqrt{\alpha^2 - \beta^2}\right)} \times \frac{K_{\lambda-\frac{1}{2}}\left(\alpha\sqrt{\delta^2 + (x-\mu)^2}\right)}{\left(\sqrt{\delta^2 + (x-\mu)^2}/a\right)^{\frac{1}{2}-\lambda}}, \tag{2.3}$$

*where*

$$K_{\lambda}(z) = \frac{1}{2}\int_0^{\infty} u^{\lambda-1} \exp\left\{-\frac{z}{2}(u + \frac{1}{u})\right\} du \tag{2.4}$$

*is the modified Bessel function of the third kind with index $\lambda$.*

*The Generalized hyperbolic Lévy process is a Lévy process $(L(t), t \geq 0)$ such that $L(1)$ follows a generalized hyperbolic distribution.*

The parameter $\mu$ controls the location of the distribution, $\alpha$ the steepness (or the fatness of the tails) of the distribution, $\beta$ the skewness, and $\delta$ is the scaling parameter. The distribution is symmetric when $\beta = 0$. The parameter $\lambda$ is identifying the subfamily within the generalized hyperbolic class. The generalized hyperbolic family is the superclass of the hyperbolic, normal inverse Gaussian, variance-gamma, normal, t-distribution.

The mean and variance of a $GH(\lambda, \alpha, \beta, \delta, \mu)$ distributed random variable are, see

Prause [1999], Lemma 1.14 :

$$\begin{aligned}
&\text{Mean} \quad \mu + \frac{\beta\delta}{\gamma}\frac{K_{\lambda+1}(\delta\gamma)}{K_\lambda(\delta\gamma)} \\
&\text{Variance} \quad \frac{\delta K_{\lambda+1}(\delta\gamma)}{\gamma K_\lambda(\delta\gamma)} + \frac{\beta^2\delta^2}{\gamma^2}\left(\frac{K_{\lambda+2}(\delta\gamma)}{K_\lambda(\delta\gamma)} - \left[\frac{K_{\lambda+1}(\delta\gamma)}{K_\lambda(\delta\gamma)}\right]\right)
\end{aligned}$$

where $\gamma = \sqrt{\alpha^2 - \beta^2}$.

The family of generalized hyperbolic distributions admits explicit moment generating function. If a random variable $X$ is distributed according to a generalized hyperbolic distribution, the moment generating function $M_{GH}(u) = \mathbb{E}[e^{uX}]$ exists for $u$ with $|\beta+u| < \alpha$ and is given by

$$M_{GH}(u) = e^{\mu u}\left(\frac{\sqrt{\alpha^2-\beta^2}}{\sqrt{\alpha^2-(\beta+u)^2}}\right)^{\lambda}\frac{K_\lambda\left(\delta\sqrt{\alpha^2-(\beta+u)^2}\right)}{K_\lambda\left(\delta\sqrt{\alpha^2-\beta^2}\right)}, \tag{2.5}$$

As a consequence, exponential moments of a generalized hyperbolic Lévy process $L(t)$ are finite, i.e. $\mathbb{E}[exp(L(t)] < \infty$. Such property is required in later pricing of weather derivatives under some martingale measure.

Now we introduce two subclasses of the generalized hyperbolic distribution, which are the hyperbolic distribution and the normal inverse Gaussian distribution. By fixing the parameter $\lambda$, they reduce the five-parameter distributions to four parameter distributions. Both distributions have been commonly applied in financial context.

Setting $\lambda = 1$ in (2.3) we obtain the hyperbolic distribution, which is first applied in finance by Eberlein and Keller [1995].

**Definition 2.8** (Hyperbolic Lévy process)**.** *The hyperbolic distribution, denoted by $H(\alpha, \beta, \mu, \delta)$, with parameters $\alpha > 0$, $|\beta| < \alpha$, $\delta > 0$ and $\mu \in \mathbb{R}$, is infinitely distribution with density function given by*

$$f_H(x;\alpha,\beta,\mu,\delta) = \frac{\sqrt{\alpha^2-\beta^2}}{2\alpha\delta K_1\left(\delta\sqrt{\alpha^2-\beta^2}\right)}\exp\left(-\alpha\sqrt{\delta^2+(x-\mu)^2}+\beta(x-\mu)\right). \tag{2.6}$$

*where $K_1(z)$ is the modified Bessel function of the third kind with index 1.*

*The Hyperbolic Lévy process is a Lévy process such that $L(1)$ follows hyperbolic distribution.*

The moment generating function of a hyperbolic distribution $H(\alpha, \beta, \mu, \delta)$ is

$$M_H(u) = e^{\mu u} \frac{\sqrt{\alpha^2 - \beta^2}}{\sqrt{\alpha^2 - (\beta + u)^2}} \frac{K_1\left(\delta\sqrt{\alpha^2 - (\beta + u)^2}\right)}{K_1\left(\delta\sqrt{\alpha^2 - \beta^2}\right)}. \tag{2.7}$$

Setting $\lambda = -\frac{1}{2}$ in (2.3), we obtain the normal inverse Gaussian (NIG) distribution which is first introduced in finance by Barndorff-Nielsen [1997].

**Definition 2.9** (Normal Inverse Gaussian process)**.** *The normal inverse Gaussian distribution, denoted by $NIG(\alpha, \beta, \mu, \delta)$, with parameters $\alpha > 0$, $|\beta| < \alpha$, $\delta > 0$ and $\mu \in \mathbb{R}$, is an infinitely divisible distribution with density function given by*

$$f_{NIG}(x; \alpha, \beta, \mu, \delta) = \frac{\alpha\delta K_1\left(\alpha\sqrt{\delta^2 + (x - \mu)^2}\right)}{\pi\sqrt{\delta^2 + (x - \mu)^2}} \exp\left(\delta\sqrt{\alpha^2 - \beta^2} + \beta(x - \mu)\right). \tag{2.8}$$

*where $K_1(z)$ is the modified Bessel function of the third kind with index 1.*

*The Normal inverse Gaussian (NIG) process is a Lévy process such that $L(1)$ is normal inverse Gaussian distributed.*

Since the Bessel function satisfies $K_{-1/2}(z) = K_{1/2}(z) = \sqrt{\pi/(2z)}e^{-z}$, see Eberlein [2009], section 4.4, $NIG(\alpha, \beta, \delta, \mu)$ distribution has more explicit form of some stylized features, which are

$$\begin{aligned}
&\text{Mean} \quad \frac{\beta\delta}{\sqrt{\alpha^2 - \beta^2}} + \mu \\
&\text{Variance} \quad \frac{\alpha^2\delta}{(\alpha^2 - \beta^2)^{3/2}} \\
&\text{Skewness} \quad \frac{3\beta}{\alpha\sqrt{\delta}(\alpha^2 - \beta^2)^{1/4}} \\
&\text{Kurtosis} \quad 3\left(1 + \frac{\alpha^2 + 4\beta^2}{\delta\alpha^2\sqrt{\delta}(\alpha^2 - \beta^2)}\right)
\end{aligned}$$

and the moment generating function of NIG distribution is

$$M_{NIG}(u) = \exp\left\{\mu u + \delta\left(\sqrt{\alpha^2 - \beta^2} - \sqrt{\alpha^2 - (\beta + u)^2}\right)\right\}. \tag{2.9}$$

If we derive $M^t_{NIG}(u)$, from (2.9), we obtain the same form of expression with parameters $t\mu$ and $t\delta$. Therefore, the NIG distributions are said to be *closed under convolution* in the two parameters $\mu$ and $\delta$, see Eberlein [2009]. From the independent and identical

increments property, the distribution of $L(t)$ is the convolution of $L(1)$, therefore for a NIG Lévy process, $L(t) \sim NIG(\alpha, \beta, \mu t, \delta t)$, for each $t \geq 0$.

We now focus back on the generalized hyperbolic distribution as the superclass. The tail behavior of the generalized hyperbolic distribution is often classified as *semi-heavy*, meaning that the tails are lighter than those non-Gaussian stable laws, but much heavier than those of the Gaussian distribution. Take the NIG distribution as an example, we see from equation (2.8) that the probability density decays exponentially like $exp(\beta x)$ as $t \to \pm\infty$. Comparatively, the normal distribution decays more rapidly like $exp(-kx^2)$ for some constant $k$. Therefore, the family of generalized hyperbolic distributions allows more probabilities of rare events compared to the normal distributions.

The generalized hyperbolic Lévy processes $L(t)$ are purely discontinuous processes with paths of *infinite variation*, see Benth and Šaltytė-Benth [2005], section 2.2. If we compare the characteristic function of $L(1)$, which is $\phi(u) = M_{GH}(iu)$, to the Lévy-Khintchine formula (2.2), we see $\sigma = 0$ and the process is a pure jump process. The infinite variation property is drawn from the Lévy measure $\ell$. For generalized hyperbolic distribution, the Lévy measure $\ell$ has explicit Lebesgue density given by

$$\ell_{GH}(dz) = |z|^{-1} e^{\beta z} \frac{1}{\pi^2} \int_0^\infty \left\{ \frac{\exp(-\sqrt{2y+\alpha^2}|z|)}{J_\lambda^2(\delta\sqrt{2y}) + Y_\lambda^2(\delta\sqrt{2y})} \frac{dy}{y} + \lambda e^{-\alpha|z|} \right\} dz, \quad \text{for } \lambda \geq 0,$$

and

$$\ell_{GH}(dz) = |z|^{-1} e^{\beta z} \frac{1}{\pi^2} \int_0^\infty \frac{\exp(-\sqrt{2y+\alpha^2}|z|)}{J_{-\lambda}^2(\delta\sqrt{2y}) + Y_{-\lambda}^2(\delta\sqrt{2y})} \frac{dy}{y} dz, \quad \text{for } \lambda < 0,$$

where $J_\lambda$ and $Y_\lambda$ are the Bessel functions of the first and second kind, respectively, with index $\lambda$. With the explicit Lévy measure, we can verify some integrability conditions imposed on Lévy measure for finance application.

Parameters of the generalized hyperbolic distribution can be estimated by maximum likelihood method. Given observations $x = (x_1, \ldots, x_n)$, the maximum likelihood estimate of the parameters $\varkappa = (\lambda, \alpha, \beta, \delta, \mu)$ can be obtained by minimizing the log-likelihood function for generalized hyperbolic distribution:

$$\begin{aligned} L(x; \varkappa) = &\log(a) + \frac{\lambda - 1/2}{2} \sum_{i=1}^{n} \log\left(\delta^2 + (x_i - \mu)^2\right) \\ &\times \log\left(K_{\lambda-1/2}\left(\alpha\sqrt{\delta^2 + (x_i - \mu)^2}\right)\right) + \beta(x_i - \mu). \end{aligned}$$

### 2.1.3 Integrals Based on Lévy Processes

In our model, the temperature dynamics is defined by a stochastic differential equation based on Lévy processes. Important results required for Lévy based stochastic calculus are presented here.

We begin with the definitions of local martingale and semimartingale. This concept is useful since the theory of stochastic integrals can be applied for calculating $\int_o^\infty H(s)dX(s)$, where the integrator $X = X(s)$ is a semimartingale and the integrand $H = H(s)$ is a predictable process.

**Definition 2.10** (Local Martingale and Semimartingale)**.** *(i) For a stochastic process $X = (X(t))$ and a stopping time $\tau$, $X^\tau := X_{t\wedge\tau}$ is called the process stopped at $\tau$. If $X$ is a process and $(\tau_n)$ an increasing sequence of stopping times, $X$ is called a local martingale if $(X^{\tau_n}) = (X_{\tau_n\wedge t})$ is a martingale. The sequence $(\tau_n)$ is called a localizing sequence.*

*(ii) An adapted, càdlàg stochastic process $X = (X(t))$ is a semimartingale if it can be decomposed as*

$$X(t) = X(0) + M(t) + C(t), \tag{2.10}$$

*where $M = M(t)$ is a local martingale and $C = C(t)$ is an adapted process of finite variation.*

From the Lévy-Itô decomposition, every Lévy process $L = L(t)$ is a semimartingale, see Applebaum [2009], page 115, Proposition 2.6.1. Then we can resort to the theory of stochastic integrals of predictable process with respect to semimartingales. For the general theory for stochastic integrals of predictable process with respect to semimartingales, we refer to He et al. [1992].

In this thesis we focus on stochastic integrals with integrable and bounded integrands. Let $L = (L(t), t \geq 0)$ be a Lévy process taking values in $\mathbb{R}$ and $f \in \mathcal{L}^2(\mathbb{R}^+)$, i.e. $f : \mathbb{R}^+ \mapsto \mathbb{R}$ satisfies

$$||f||_2 = \left[\int_0^\infty |f(s)|^2 ds\right]^{1/2} < \infty.$$

Note that we use $\mathcal{L}^2$ to denote the space with the above norm $||\cdot||_2$. We consider the *Wiener-Lévy integrals* $Y = (Y(t), t \geq 0)$, where each

$$Y(t) = \int_0^t f(s)dL(s). \tag{2.11}$$

First we have that the process $Y = Y(t)$ has independent increments, see Applebaum [2009], page 214, Lemma 4.3.12, which is

**Lemma 2.11.** *For each $0 \leq s < t < \infty$, $Y(t) - Y(s)$ is independent of $\mathcal{F}_s$.*

Typically, if $L$ in equation (2.11) is a standard Brownian motion $B = (B(t), t \geq 0)$, we have that for each $t \geq 0$,

$$Y(t) = \int_0^t f(s)dB(s) \sim N\left(0, \int_0^t |f(s)|^2 ds\right). \tag{2.12}$$

For the differential-integral calculus based on Lévy process, we need an important tool known as the Itô's Lemma. Here we present the multidimensional Itô's Formula. For more information, we refer to Watanabe and Ikeda [1981].

**Theorem 2.12** (Itô's Lemma for semimartingales)**.** *Let $X^1, \ldots, X^d$ be semimartingales, and $X(t) = (X^1(t), \ldots, X^d(t))$. Each and $f$ is a function on $\mathbb{R}^d$ with continuous partial derivatives of the first and second orders. Then $f(X(t)) = (f(X^1(t), \ldots, X^d(t)))_{t\geq 0}$ is a semimartingale and*

$$\begin{aligned} f(X(t)) =& f(X(0)) + \sum_{j=1}^{d} \int_0^t f_j(X(s-)dX^i(s) + \frac{1}{2}\sum_{i,j=1}^{d} \int_0^t f_{i,j}(X(s-))d\langle (X^i)^c, (X^j)^c\rangle(s) \\ &+ \sum_{0<s\leq t} \left( f(X(s)) - f(X(s-)) - \sum_{i=j}^{d} f(X(s-))\Delta X^j(s) \right) \end{aligned}$$

*where $f_i = \frac{\partial f}{\partial x_j}$, $f_{i,j} = \frac{\partial^2 f}{\partial x_i \partial x_j}$, $X(t-)$ is the left limit in $t$ of process $X$, $\Delta X(t) = X(t) - X(t-)$ , $\langle X, Y\rangle$ is the predictable quadratic covariance of $X$ and $Y$ (see He et al. [1992], page 186), and $X^c$ is the continuous martingale part of $X$.*

## 2.2 Arbitrage Theory

The arbitrage theory for pricing of derivatives is often applied after establishing a model for the dynamics of the underlying asset. Later we will price the temperature derivatives using no arbitrage theory. As a reference, we present some basic concepts and results based on Bingham and Kiesel [2004] in this section. The main focus is on the concept of arbitrage-free and completeness, and their relations to martingale measure.

A probability space $(\Omega, \mathcal{F}, P)$ with a filtration $\mathbb{F} = (\mathcal{F}_t)_{0\leq t\leq \tau_{max}}$ satisfying the usual hypothesis is assumed given to model random events in the market. Here $\tau_{max} < \infty$ denotes the fixed maximal time up to which trades are allowed and $\mathcal{F}_{\tau_{max}} = \mathcal{F}$. The market model consists of $d + 1$ basic assets whose price processes are modelled by stochastic processes $S_0, \ldots, S_d$ defined on $(\Omega, \mathcal{F}, \mathbb{F}, P)$. The continuously compounded risk-free interest rate,

$r$, is assumed to be constant. Let $S_0(t) = R(t)$ be the risk-free bond given by $R(t) = e^{rt}$. $R(t)$ is a *numéraire*, i.e. an almost surely strictly positive price process for $t \in [0, \tau_{max}]$. The *discounted price process* is defined as

$$\tilde{S}(t) = \frac{S(t)}{R(t)} = \left(1, \tilde{S}_1(t), \ldots, \tilde{S}_d(t)\right), \quad t \in [0, \tau_{max}],$$

with $\tilde{S}_i(t) = \frac{S_i(t)}{R(t)}$, for $i = 1, \ldots, d$.

An $\mathbb{R}^{d+1}$-valued predictable process

$$\vartheta(t) = (\vartheta_0(t), \ldots, \vartheta_d(t)), \quad t \in [0, \tau_{max}],$$

a *trading strategy* (or *dynamic portfolio process*), if $\vartheta_i$ is sufficiently integrable such that $\int_0^t \vartheta_i(s) dS_i(s)$ is well-defined. Here $\vartheta_i(t)$ denotes the number of shares of asset $i$ held in the portfolio at time $t-$, which denotes the left limit of $t$. Some definitions on trading strategies are required before introducing arbitrage opportunities.

**Definition 2.13** (Value process, gains process and self-financing)**.** *(i) The value (or wealth) process of the trading strategy $\vartheta$ is defined as*

$$V_\vartheta(t) := \vartheta(t)^\top S(t) = \sum_{i=0}^{d} \vartheta_i(t) S_i(t), \quad t \in [0, \tau_{max}].$$

*(ii) The gains process based on trading strategy $\vartheta$ is defined as*

$$G_\vartheta(t) := \int_0^t \vartheta(t)^\top dS(t). = \sum_{i=0}^{d} \int_0^t \vartheta_i(t) dS_i(t).$$

*(iii) A trading strategy $\vartheta$ is called self-financing if the wealth process $V_\vartheta$ satisfies*

$$V_\vartheta(t) = V_\vartheta(0) + G_\vartheta(t) \quad \textit{for all } t \in [0, \tau_{max}].$$

**Definition 2.14** (Arbitrage opportunity)**.** *A self-financing trading strategy $\vartheta$ is called an arbitrage opportunity is the wealth process $V_\vartheta$ satisfies the following set of conditions:*

$$V_\vartheta(0) = 0, \quad P(V_\vartheta(\tau_{max}) \geq 0) = 1, \quad \textit{and} \quad P(V_\vartheta(\tau_{max}) > 0) > 0.$$

By definition, an arbitrage opportunity allows a risk-free profit. The arbitrage theory is based on the assumption that the market does not allow any arbitrage opportunity. In an

efficient market, such no-arbitrage assumption should be true. The concept of equivalent martingale measures, given in the following definition, plays an important role in arbitrage theory.

**Definition 2.15** (Equivalent martingale measure)**.** *A probability measure $Q$ defined on $(\Omega, \mathcal{F})$ is called an equivalent martingale measure (EMM) if*

*(i) $Q$ is equivalent to $P$, denoted as $Q \sim P$,*

*(ii) the discounted price processes $\tilde{S}_i$, $i = 1, \ldots, d$, are $Q$-martingales.*

*We denote the set of martingale measures by $\mathcal{Q}$.*

Discussion on the link between the arbitrage and EMM follows. First some restrictions are imposed on trading strategies from following Definition.

**Definition 2.16.** *A self-financing strategy $\vartheta$ is called tame (relative to the numéraire) if*

$$\tilde{V}_\vartheta \geq 0, \quad \text{for } t \in [0, \tau_{max}].$$

*The set of tame strategies is denoted by $\Upsilon$.*

**Theorem 2.17.** *Assume that there exists an EMM, then the market contains no arbitrage opportunities in $\Upsilon$.*

The converse statement of Theorem 2.17 , i.e. no arbitrage implying the existence of an EMM, is basically true as well. One requires a more technical and stronger definition of no arbitrage, known as No free lunch with vanishing risk (NFLVR), see Bingham and Kiesel [2004], page 235, Definition 6.1.6 and Theorem 6.1.2. In general, the equivalence between no arbitrage and existence of equivalent martingale measure is known as the *First Fundamental Theorem of Asset Pricing*.

Now we turn to the problem of pricing a financial derivative under the no-arbitrage assumption. Under the assumption, there exists an EMM $Q$. The cash flow of a derivative is described via the concept of contingent claims, which is an $\mathcal{F}$-measurable random variable $X$ such that $\tilde{X} := X/S_0(\tau_{max})$ satisfies $\mathbb{E}_Q[\int_0^{\tau_{max}} \tilde{X}(t)dt] < \infty$. In order to price or hedge the contingent claim, we attempt to construct a trading strategies which replicate the its payoff. Mathematically, this is described in the following Definition.

**Definition 2.18.** *(i) A self-financing strategy $\vartheta$ is called $Q$-admissible if the relative gains process*

$$\tilde{G}_\vartheta = \int_0^t \vartheta(u)^\top d\tilde{S}(u)$$

*is a $Q$-martingale.*

*(ii) A contingent claim $X$ is called attainable if there exists at least one admissible trading strategy such that*

$$V_\vartheta(\tau_{max}) = X.$$

*We call such a trading strategy $\vartheta$ a replicating strategy for $X$.*

From the definition, if the claim $X$ is attainable, we can construct a portfolio $\vartheta$ which returns the same payoff as $X$ at time $\tau_{max}$. Holding the portfolio is equivalent to holding the contingent claim $X$, and hence they their value should be the same. From the below Theorem, we can obtain the arbitrage-free price process $\Pi_X(t)$ of an contingent claim $X$.

**Theorem 2.19** (Risk-neutral Valuation Formula)**.** *The arbitrage price process of any attainable claim $X$ is given by the risk neutral formula*

$$\Pi_X(t) = e^{-(\tau_{max}-t)}\mathbb{E}_Q\left[X|\mathcal{F}_t\right], \quad \textit{for all } t \in [0, \tau_{max}]. \tag{2.13}$$

In order to investigate the uniqueness of the EMM $Q$, the definition of a complete market is introduce.

**Definition 2.20.** *The financial market model is said to be complete if any contingent claim is attainable.*

Completeness of a market implies that we have sufficient securities in the market so that any new one can be replicated by the existing ones. The market is incomplete if we do not have sufficient securities to construct a replicating portfolio for all contingent claim. This is often the case when the asset price processes have jumps and stochastic volatility. As posed in Staum [2007], in a market where jumps of all sized are possible, one need, for example, vanilla European options of all strikes and maturities, to complete the market. In market For weather derivatives, the underlying asset is not traded, this constitutes another type of insufficiency.

Completeness of the market is linked to uniques of EMM in the following Theorem.

**Theorem 2.21** (Second Fundamental Theorem of Asset Pricing)**.** *A market model that admits at least one EMM $Q$ is complete if and only if $Q$ is unique.*

If the market is arbitrage-free and complete, then there exists only one unique EMM $Q$ such that the price of a contingent claim $X$ given by (2.13) is unique. In the case where the market is arbitrage-free but is not complete, the market admits more than one

equivalent martingale measures, then the price is in the interval

$$\left( \inf_{Q \in \mathcal{Q}} \mathbb{E}_Q[\tilde{X}], \sup_{Q \in \mathcal{Q}} \mathbb{E}_Q[\tilde{X}] \right).$$

where $\mathcal{Q}$ is the set of all EMMs, $\tilde{X}$ is the discounted payoff, see Delbaen and Schachermayer [2006], page 24, Theorem 2.4.1.

The problem is now to select the most suitable EMM $Q$ in $\mathcal{Q}$ and then to value all claims as the expected discounted value with respect to $Q$. In financial application, we have several kinds of candidates for constructions of equivalent martingale. The Esscher transformation method we introduced in next section is one of the well known techniques to obtain a manageable class of EMMs.

## 2.3 Esscher Transform

In financial application, one usually has to change the physical probability measure $P$ to an EMM $Q$, known as the risk neutral measure, such that the discounted price process of the underlying is a martingale under $Q$. Under the setting of Brownian motion, we have Girsanov transform to serve the purpose of measure transformation. Under the setting of Lévy process, one can resort to Esscher transform. The Esscher transform is introduced by Esscher [1932] to study risk theory and then applied by Gerber and Shiu [1994] to option pricing.

First we review the Girsanov's theorem.

**Theorem 2.22** (Girsanov's theorem)**.** *Let $B$ be a Brownian Motion on $(\Omega, \mathcal{F}, \mathbb{F}, P)$. With $\theta(t)$ being a real-valued, and bounded and piecewise continuous function, let*

$$Z^{\theta}_{BM}(t) = \exp\left( \int_0^t \theta(s) dL(s) - \frac{1}{2} \int_0^t ||\theta(s)||^2 ds \right).$$

*Define*

$$B^{\theta}(t) := B(t) - \int_0^t \theta(u) du.$$

*We define a equivalent probability measure $Q^{\theta}$ by the Radonr-Nikodym derivative*

$$\frac{dQ^{\theta}}{dP} | \mathcal{F}_t = Z^{\theta}_{BM}(t),$$

*then the process $B^{\theta}(t)$ is a Brownian motion with respect to $Q^{\theta}$.*

As we have seen, the Girsanov transform preserves the normality of the distribution of the Brownian motion. The Esscher transform is a generalization of the Girsanov transform of Brownian motion to jump processes. The following theorem from Alexandridis and Zapranis [2013], Theorem 2.7, states its original definition.

**Theorem 2.23.** *The Esscher Transform. Let $f(x)$ be a probability density; then its Esscher transform is defined as*

$$f(x,\theta) = \frac{e^{\theta x} f(x)}{\int_{-\infty}^{+\infty} e^{\theta x} f(x) dx} \tag{2.14}$$

*More generally, if $P$ is a probability measure, then the Esscher transform of $P$ is a new probability measure $Q^{\theta}$ which has density*

$$\frac{e^{\theta x} f(x)}{\int_{-\infty}^{+\infty} e^{\theta x} f(x) dP(x)} \tag{2.15}$$

*with respect to $P$.*

The Esscher transform preserves the distributional properties of jump processes, so that we still have access to their characteristics, as posed in Benth et al. [2008]. Moreover, the characteristics of the jump process is altered by the parameter $\theta$, while the independent increment property is yet preserved, see Benth et al. [2008]. This is an useful technique in the context of Lévy processes as we will see later when it is applied on pricing derivatives.

## 2.4 Basics for Time Series Analysis

In this section, we introduce the discrete-time autoregressive moving average (ARMA) process, based on Brockwell and Davis [2009]. We also include some statistical tests in the context of time series analysis for further use of model diagnostic.

### 2.4.1 Discrete-time Autoregressive and Moving Average Processes

The discrete-time ARMA process is an important class of stationary stochastic process of the discrete form $\{X(t), t \in \mathbb{Z}\}$. It can be viewed as a discrete counterpart of the continuous-time autoregressive moving average (CARMA) process which will be introduced in next section. It is also relevant for further data analysis.

**Definition 2.24** (The Autocovariance Function)**.** *If $(X(t), t \in \mathcal{T})$ is a process such that $Var(X(t)) < \infty$ for each $t \in \mathcal{T}$, then the autocovariance function, $\gamma_X(\cdot,\cdot)$, of $\{X(t)\}$ is defined by*

$$\gamma_X(s,t) = Cov(X(s), X(t)) = \mathbb{E}\left[(X(s) - \mathbb{E}X(s))\,(X(t) - \mathbb{E}X(t))\right], \quad s,t \in \mathcal{T},$$

*where $\mathbb{E}(\cdot)$ denotes the expectation.*

**Definition 2.25** (The Autocorrelation Function)**.** *The autocorrelation function (ACF), $\rho_X(\cdot,\cdot)$, of $(X(t), t \in \mathcal{T})$ is defined by*

$$\rho_X(s,t) = Corr(X_s, X_t) = \frac{\gamma_X(s,t)}{\sqrt{\gamma_X(r,r)\gamma_X(s,s)}}, \quad s,t \in \mathcal{T}.$$

**Definition 2.26** (Stationarity)**.** *The time series $\{X(t), t \in \mathbb{Z}\}$ is said to be stationary if*

*(i)* $\mathbb{E}|X_t|^2 < \infty$ *for all* $t \in \mathbb{Z}$,

*(ii)* $\mathbb{E}X_t = m$ *for all* $t \in \mathbb{Z}$,

*(iii)* $\gamma_X(s,t) = \gamma_X(s+r, t+r)$ *for all* $s,t,r \in \mathbb{Z}$.

*Such stationarity is referred to as weak stationarity, covariance stationarity or second-order stationarity.*

**Definition 2.27** (White Noise)**.** *The process $\{Z(t)\}$ is said to be white noise with mean 0 and variance $\sigma^2$, denoted as*

$$\{Z(t)\} \sim WN(0, \sigma^2),$$

*if and only if $\{Z(t)\}$ has zero mean and covariance function*

$$\gamma_Z(h) = \begin{cases} \sigma^2 & h = 0, \\ 0 & h \neq 0. \end{cases}$$

**Definition 2.28** (ARMA($p$,$q$) Process)**.** *The process $\{X(t), t \in \mathbb{Z}\}$ is said to be an ARMA(p,q) process if $\{X(t)\}$ is stationary and for every $t$,*

$$X(t) - b_1X(t-1) - \cdots - b_pX(t-p) = Z(t) + a_1Z(t-1) + \cdots + a_qZ(t-q), \quad (2.16)$$

*where $\{Z_t\} \sim WN(0, \sigma^2)$. We say that $\{X(t)\}$ is an ARMA(p,q) process with mean $\mu$ if $\{X(t) - \mu\}$ is an ARMA(p,q) process.*

We can rewrite equation (2.16) as

$$b(\mathrm{B})X(t) = a(\mathrm{B})Z(t), \tag{2.17}$$

where B is the backward shift operator, i.e. $\mathrm{B}^j X(t) = X(t-j)$, for each $j \in \mathbb{Z}$; $a$ and $b$ are the $p$th and $q$th degree polynomials

$$a(z) := 1 - a_1 z - \cdots - a_p z^p,$$

$$b(z) := 1 + b_1 z + \cdots + b_q z^q.$$

Following are two special cases of the ARMA process. Setting $a(z) \equiv 1$ in equation (2.17), we obtain the autoregressive (AR) process. Setting $b(z) \equiv 1$ in equation (2.17), we obtain the moving average (MA) process.

**Definition 2.29** (The AR($p$) process)**.** *The process $(X(t), t \in \mathbb{Z})$ is said to be an AR(p) process if $\{X(t)\}$ is stationary and for every $t$,*

$$X(t) - b_1 X(t-1) - \cdots - b_p X(t-p) = Z(t) \tag{2.18}$$

*where $\{Z_t\} \sim WN(0, \sigma^2)$.*

**Definition 2.30** (The MA($q$) process)**.** *The process $(X(t), t \in \mathbb{Z})$ is said to be an MA(q) process if $\{X(t)\}$ is stationary and for every $t$,*

$$X(t) = Z(t) + a_1 Z(t-1) + \cdots + a_q Z(t-q), \tag{2.19}$$

*where $\{Z_t\} \sim WN(0, \sigma^2)$.*

Now we present the definitions of the autocorrelation function (ACF) and partial autocorrelation functions (PACF) of stationary process defined in Definition 2.26. In practice, we often estimate the ACF and PACF of a stationary time series $\{X(t)\}$ from observations $\{x_1, x_2, \ldots, x_n\}$ in order to explore the dependence structure of $\{X(t)\}$. For a stationary process defined in Definition 2.26, the covariance of $X(s)$ and $X(t)$ only depends on $|s-t|$, because $\gamma_X(s,t) = \gamma_X(s-t, 0)$, for $s, t \in \mathbb{Z}$. Therefore, we define the autocovariance function of a stationary time series as

$$\gamma_X(h) := \gamma_X(h, 0) = Cov(X(t+h), X(t)), \quad t, h \in \mathbb{Z}.$$

where $h$ is called the *lag*. Analogously, the autocorrelation function (ACF) of a stationary time series at lag $h$ is defined as

$$\rho_X(h) := \frac{\rho_X(h)}{\rho_X(0)} = Corr(X(t+h), X(t)), \quad t, h \in \mathbb{Z}.$$

In order to explore more about the dependence structure of a stationary process, we introduce the partial autocorrelation function.

**Definition 2.31** (Partial Autocorrelation Function)**.** *The partial autocorrelation function (PACF) of a stationary time series is defined by*

$$\alpha(1) = Corr(X_2, X_1) = \rho(1)$$

*and*

$$\alpha(k) = Corr\left(X_{k+1} - \mathrm{P}_{\bar{sp}\{1,X_2,\ldots,X_k\}}X_{k+1}, X_1 - \mathrm{P}_{\bar{sp}\{1,X_2,\ldots,X_k\}}X_1\right), \quad k \geq 2,$$

*where* $\mathrm{P}_{\bar{sp}\{1,X_2,\ldots,X_k\}}X$ *denotes the projection of* $X$ *onto space spanned by* $\{1, X_2, \ldots, X_k\}$.

The partial autocorrelation $\alpha(k)$, $k \geq 2$, estimates the adjusted correlation of $X_{k+1}$ and $X_1$, by removing the influence of $1, X_2, \ldots, X_k$.

Table 2.1 summarizes the general behavior of ACFs and PACFs of AR, MA and ARMA processes. It helps to detect the order $p$ and $q$ of ARMA process in practice.

Table 2.1: General behavior of ACFs and PACFs of AR, MA ARMA processes

| | AR($p$) | MA($q$) | ARMA($p$,$q$) |
|---|---|---|---|
| ACF | Tails off | Cuts off after lag $q$ | Tails off |
| PACF | Cuts off after lag $p$ | Tails off | Tails off |

The estimation of $\gamma(h)$ and $\rho(h)$ are sample autocovariance function and sample autocorrelation function defined as following.

**Definition 2.32** (Sample Autocovariance Function)**.** *The sample autocovariance function of observations* $\{x_1, x_2, \ldots, x_n\}$ *is defined as*

$$\hat{\gamma}(h) := \frac{1}{n}\sum_{j=1}^{n-h}(x_{j+h} - \bar{x})(x_j - \bar{x}), \quad 0 \leq < n, \tag{2.20}$$

*and* $\hat{\gamma}(h) = \hat{\gamma}(-h)$, $-n < h \leq 0$, *where* $\bar{x} = \frac{1}{n}\sum_{j=1}^{n} x_j$.

**Definition 2.33** (Sample Autocorrelation Function (ACF)). *The sample ACF of observations* $\{x_1, x_2, \ldots, x_n\}$ *is defined as*

$$\hat{\rho}(h) := \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}, \quad |h| < n. \tag{2.21}$$

Analogously, we have *sample PACF* for observations $\{x_1, x_2, \ldots, x_n\}$. For more information, we refer to Brockwell and Davis [2009].

### 2.4.2 Statistical Tests for Model Diagnostics

Here we introduce some model selection criteria and statistical tests we will come across in fitting the model.

#### Information Criteria

Given a finite set of models, we need some general model selection criteria in order to select the most suitable one.

The *Akaike information criterion (AIC)*, introduced by Akaike [1992], is one of the most common means for model selection. Given observation $\{x_i, i = 1, \ldots, n\}$, the AIC is a value defined as

$$AIC := 2k - 2\ln(Lik) \tag{2.22}$$

where $k$ is the number of parameters to be estimated in the model, $Lik$ denotes the maximized value of the likelihood function for the estimated model. The larger the value of the likelihood function, the smaller the AIC value is. The term $2k$ acts as a penalty to avoid over fitting, hence it provides a trade-off between the goodness of fit and the complexity of the model. Consequently, the model with the minimum AIC value is preferred.

The *Bayesian information criterion (BIC)* or *Schwartz criterion*, introduced by Schwarz [1978], is another widely used criterion based on the likelihood function. The formula for BIC is constructed similarly to AIC, only with a larger penalty term

$$BIC := k\ln(n) - 2\ln(Lik) \tag{2.23}$$

where $k$ is the number of parameters to be estimated in the model; $n$ is the number of observations; $Lik$ is the maximized value of the likelihood function of the estimated model. Similar to the case of AIC, the model with the minimum BIC value is preferred.

**Test of Stationarity**

Normally time series data has seasonal mean and is not stationary. We usually remove the trend and seasonality in the data first, and then work on stationary data. Test of stationarity is then required. We begin with the concept of unit root which is closely related to stationarity. Consider a process $\{X(t), t \in \mathbb{Z}\}$ and suppose that it is an AR($p$) process

$$X(t) = \phi_1 X(t-1) + \cdots + \phi_p X(t-p) + \epsilon(t),$$

where $\{\epsilon(t)\} \sim WN(0, \sigma^2)$. If the characteristic equation

$$m^p - \phi_1 m^{p-1} - \cdots - \phi_p = 0$$

has a root $m = 1$, then the process $X(t)$ has a *unit root*. Such a process is non-stationary. Here we introduce two unit root tests for examining the stationarity of time series.

The *Augmented Dickey Fuller (ADF) test* is one commonly used unit root test. Dickey and Fuller [1979] develops a procedure for testing whether a variable has a unit root. The ADF test is an augmented version of the Dickey-Fuller test. The ADF test is performed on the model of following form:

$$\Delta X(t) = \rho X(t-1) + \xi_1 \Delta X(t-1) + \cdots + \xi_{p-1} \Delta X(t-p+1) + \epsilon(t),$$

where $\Delta$ is the difference operator, $\rho = \phi_1 + \cdots + \phi_p - 1$ and $\xi_i = -(\phi_{i+1} + \cdots + \phi_p)$. The augmenting lags $p$ need to be determined before applying the test. It can chosen by minimizing information criteria such as AIC or BIC, or dropping lags from high to low order based on statistical significance of corresponding estimated coefficients. The hypotheses of the ADF test are

$$\begin{aligned} H_0: &\quad \rho = 0 \quad \text{(the process has a unit root, and is non-stationary)} \\ H_1: &\quad \rho < 0 \quad \text{(the process is stationary)} \end{aligned}$$

We reject the null hypothesis if the test statistic is smaller than the critical value at certain significance level. For more details we refer to Dickey and Fuller [1979].

The *Kwiatkowski Phillips Schmidt Shin (KPSS)* test, developed by Kwiatkowski et al. [1992], tests the null hypothesis that the time series is stationary versus the alternative that the time series has a unit root (non-stationary). The regression model with a time

trend can be written as

$$X(t) = c + \mu t + k \sum_{i=1}^{t} \zeta_i + \varepsilon(t)$$

with stationary $\varepsilon(t)$ and $\zeta_i \sim iid(0, 1)$. The hypotheses of KPSS test are

$$\begin{aligned} H_0 : &\quad k = 0, \text{ the process is trend-stationary (level-stationary if } \mu = 0) \\ H_1 : &\quad k \neq 0, \text{ the process is not (trend or level) stationary} \end{aligned}$$

We reject the null hypothesis if the KPSS test statistic is larger than the critical value at certain significance level. For more details we refer to Kwiatkowski et al. [1992].

### Test for Correlation and Independence

To detect serial dependence we can perform the *Ljung-Box test*, developed by Ljung and Box [1978]. The Ljung-Box test tests the hypotheses

$$\begin{aligned} H_0 &: \text{The data are independently distributed} \\ H_1 &: \text{The data are not independently distributed} \end{aligned}$$

The Ljung-Box test jointly consider the sample autocorrelations of the first $m$ lags with the test statistic:

$$\tilde{Q}_m = n(n+2) \sum_{k=1}^{m} \frac{\hat{\rho}_k^2}{n-k} \tag{2.24}$$

where $n$ is the sample size, $\hat{\rho}_k$ is the sample autocorrelation at lag $k$ and $m$ is the number of lags being tested. Under the null hypothesis $\tilde{Q}_m$ is asymptotically $\chi^2$ distributed with $m - s$ degrees of freedom, with $s$ being the number of parameters estimated in the model, see Ljung and Box [1978]. For significance level $\alpha$, the null hypothesis is rejected if

$$\tilde{Q}_m > \chi^2_{m-s,1-\alpha},$$

where $\chi^2_{m-s,1-\alpha}$ is the $100(1-\alpha)\%$-quantile of the $\chi^2$ distribution with $m - s$ degrees of freedom.

In our data analysis, the Ljung-Box test will be applied on residuals of a fitted ARMA model in order to make sure that serial dependence is removed form the original time series. For residuals of an ARMA($p$,$q$) process, the number of parameters to be estimated is $s = p + q$, hence $\chi^2$ distribution with $m - p - q$ degrees of freedom should be used.

**Test of Distributional Assumptions**

In order to evaluate how well the model fits a set of observations, we need so called *goodness-of-fit tests*. Here we introduce some tests we will apply later in data analysis.

The *Jarque-Bera (J-B) test* is a two-sided goodness-of-fit test of whether a sample of data is drawn from a normal distribution. The hypothesis is:

$$H_0 = \text{The sample data comes from a normal distribution}$$
$$H_1 = \text{The sample data does not come from a normal distribution}$$

The test statistic is given by

$$JB = \frac{n}{6}\left(Skewness^2 + \frac{(Kurtosis - 3)^2}{4}\right) \tag{2.25}$$

where $n$ is the sample size, $Skewness$ denotes the sample skewness, and $Kurtosis$ denotes the sample kurtosis. The J-B test performs by testing whether the skewness and kurtosis of the sample data match normal distribution. Under the null hypothesis, the JB statistic asymptotically has a $\chi^2$ distribution with two degrees of freedom.

The (one-sample) *Kolmogorov-Smirnov (K-S) test* is a nonparametric test that can be used to compare empirical distribution of sample data with a reference distribution. The hypothesis of the K-S test is given by

$$H_0 = \text{The data have the hypothesized distribution}$$
$$H_1 = \text{The data do not have the hypothesized distribution}$$

The K-S statistic is *K-S distance* given as

$$KS = \sqrt{n}\sup_x |F(x) - F_n(x)|, \tag{2.26}$$

where $n$ is the sample size, $F_n(x)$ is the empirical *cumulative density function (CDF)* and $F(x)$ is the estimated CDF. The K-S distance estimate the distance between the empirical distribution of the residuals and the theoretical one of the assumed distribution. If $F(x)$ is continuous, KS converges to the Kolmogorov distribution under the null hypothesis. Hence the critical value can be found under Kolmogorov distribution and does not depend on $F(x)$.

The *Anderson Darling (A-D) test*, developed by Anderson and Darling [1952], is also a nonparametric test of whether a given sample of data comes from a given probability

distribution. The test statistic is given by

$$AD = \sup_x \frac{|F(x) - F_n(x)|}{\sqrt{F(x)(1 - F(x))}} \tag{2.27}$$

The A-D statistics gives more weight to the tails of the distribution compared to the K-S statistics. The critical value is currently available for normal, lognormal, exponential, Weibull, extreme value type I and logistic distribution. For both K-S and A-D tests, smaller test statistic indicates closer distance between the empirical distribution and the reference distribution, which implies a better fit.

An graphical method for comparing two probability distribution is *Quantile-Quantile (Q-Q) plot*, which is a plot of the quantiles of two distributions against each other. A 45-degree reference line is also plotted. If the two sets come from a population with the same distribution, the points should fall approximately along this reference line. The greater the departure from this reference line, the greater the evidence for the conclusion that the two data sets have come from populations with different distributions.

# Chapter 3

# Continuous-time Model for Temperature

Here we introduce the Lévy driven continuous-time autoregressive (CAR) process and specify the CAR model we use to model daily average temperature (DAT).

Definition and properties of CAR process are presented in section 1. Section 2 introduces the link between CAR process and its discrete counterpart. The CAR model for temperature is described in section 3.

## 3.1 Continuous-time Autoregressive Process

Here we introduce the continuous-time autoregressive (CAR) process with seasonal variance. The class of CAR processes is first studied by Phillips [1959]. It is a subclass of the more general continuous-time autoregressive moving-average (CARMA) processes. Lévy processes are included in the definition of CARMA process recently in order to obtain a flexible class of continuous-time stationary processes allowing (semi-)heavy tail behaviors.

First we present the definition of the Lévy driven CAR process with constant volatility based on Brockwell [2001]. Here we consider Lévy processes satisfying $\mathbb{E}[L(1)]^2 < \infty$. Moreover, we assume that $L$ is scaled so that $Var(L(1)) = 1$, then $Var(L(t)) = t$ for all $t \geq 0$ and there exists a real constant $\mu$ such that $\mathbb{E}[L(t)] = \mu t$ for all $t \geq 0$. We call such Lévy process $L$ a *standardized second-order Lévy processes*. Without further mentions, we limit the scope to the standardized second-order Lévy processes in this chapter. The Lévy-driven CAR process is a special case of the Lévy-driven continuous-time ARMA process defined in Brockwell [2001].

**Definition 3.1** (Lévy driven CAR process)**.** *A Lévy driven continuous-time autoregressive (CAR) process of order p is defined via the state-space representation of the formal*

*equation*

$$\alpha(\mathrm{D})Y(t) = b\mathrm{D}L(t), \tag{3.1}$$

*where* $b > 0$ *is a scale parameter,* $(L(t), t \geq 0)$ *is a Lévy process,* D *denotes the differentiation with respect to* $t$ *and*

$$a(z) = z^p + a_1 z^{p-1} + \cdots + a_p.$$

*Since the derivative* $\mathrm{D}L(t)$ *does not exist in the usual sense, equation (3.1) is interpreted as being equivalent to the following observation and state equations*

$$Y(t) = \mathbf{b}^\top \mathbf{X}(t), \tag{3.2}$$

*and the state vector* $\mathbf{X}(t) = (X_1(t), \ldots, X_p(t))^\top$ *satisfies the Itô equation*

$$d\mathbf{X}(t) = \mathbf{A}\mathbf{X}(t)dt + \mathbf{e}_p dL(t), \tag{3.3}$$

*where* $\mathbf{b} = (b, 0, \ldots, 0)^\top$, $\mathbf{A}$ *is a* $p \times p$*-matrix of the form:*

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 & \ldots & 0 & 0 \\ 0 & 0 & 1 & \ldots & 0 & 0 \\ \vdots & & \ddots & & 1 & \vdots \\ 0 & \ldots & \ldots & 0 & 0 & 1 \\ -a_p & -a_{p-1} & \ldots & & -a_2 & -a_1 \end{bmatrix}, \tag{3.4}$$

*with positive constants* $a_1, \ldots, a_p$; $\mathbf{e}_k$ *denotes the* $k$*-th unit vector in* $\mathbb{R}^p$ *for* $k = 1, \ldots, p$; $\mathbf{b}^\top$ *denotes the transpose of* $\mathbf{b}$.

If $p = 1$, $b = 1$ and $L$ is a Brownian motion $B$, the CAR(1) process $Y(t) = X(t)$ is a *Gaussian Ornstein-Uhlenbeck process.* $X(t)$ defined by the state equation (3.3), known as the *Langevin's equation*,

$$dX(t) = -a_1 X(t)dt + dB(t), \tag{3.5}$$

has explicit solution

$$X(t) = e^{-a_1 t}X(0) + \int_0^t e^{-a_1(t-u)}dB(u), \tag{3.6}$$

for each $t \geq 0$. $X$ defined by (3.6) satisfies (3.5) with any $a_1$ and initial value $X(0)$. Moreover, it is the unique, strong Markov solution to (3.5), cf. Protter [2004], page 297. The process $X$ is mean reverting to 0. From (3.5), when $X$ is above 0 at some time, then the coefficient of the $dt$ drift term is negative (since $a_1 > 0$), so $X$ will tend to move

downwards immediately after. The reverse holds if $X$ is lower than zero.

Now we extend the Lévy driven CAR process defined above by incorporating seasonal volatility in the state equation (3.3), as introduced in Benth et al. [2007]. Consider the stochastic process $\mathbf{X}(t)$ in $\mathbb{R}^p$ for $p \geq 1$ defined by the vectorial Ornstein-Uhlenbeck equation

$$d\mathbf{X}(t) = \mathbf{A}\mathbf{X}(t)dt + \mathbf{e}_p\sigma(t)dL(t), \tag{3.7}$$

where $\sigma(t) > 0$ is a real-valued function and square integrable over any finite time interval, $L$ is a Lévy process and $\mathbf{A}$ is $p \times p$-matrix as (3.4) with positive constants $a_1, \ldots, a_p$.

In the case when $L$ is a standard Brownian motion $B$, the solution to the stochastic differential equation (3.7) exists and has explicit form

$$\mathbf{X}(t) = e^{\mathbf{A}t}\mathbf{X}(0) + \int_0^t \sigma(u)e^{\mathbf{A}(t-u)}\mathbf{e}_p dB(u), \quad \text{for } t \geq 0, \tag{3.8}$$

where

$$e^{\mathbf{A}t} := \sum_{k=0}^{\infty} \frac{t^k}{k!}A^k.$$

This is obtained by applying multi-dimensional Itô's Lemma, see Øksendal [2003], so that

$$d\left(e^{-\mathbf{A}t}\mathbf{X}(t)\right) = e^{-\mathbf{A}t}\left(d\mathbf{X}(t) - \mathbf{A}\mathbf{X}(t)dt\right) = e^{-\mathbf{A}t}\sigma(t)\mathbf{e}dB(t).$$

As discussed in Brockwell [2009], the integral in (3.8) is defined as the $\mathcal{L}^2$ limit of approximating Riemann-Stieltjes sums $S_n$ corresponding to the partition of the interval $[0, t]$ by points $\{k/2^n, k \in \mathbb{Z}, 0 \leq k < 2^n t\}$ and $t$. By using Itô's Lemma for semimartingale, we obtain the solution to (3.7), that is

$$\mathbf{X}(t) = e^{\mathbf{A}t}\mathbf{X}(0) + \int_0^t e^{\mathbf{A}(t-u)}\mathbf{e}_p\sigma(u)dL(u), \quad \text{for } t \geq 0, \tag{3.9}$$

The integral in (3.9) is defined in the same way as above for second-order Lévy processes. The continuous differentiability of the integrand in (3.9) implies that the sequence $\{S_n\}$ converges geometrically rapidly in $\mathcal{L}^2$ and hence almost surely to the same limit. In fact, the integral in (3.9) with deterministic and continuously differential integrand is a special case of integration with respect to semimartingale. From (3.9), we can also write

$$\mathbf{X}(s) = e^{\mathbf{A}(s-t)}\mathbf{X}(t) + \int_t^s e^{\mathbf{A}(s-u)}\mathbf{e}_p\sigma(u)dL(u), \quad \text{for } s \geq t \geq 0. \tag{3.10}$$

From (3.10) we see that $(\mathbf{X}(t), t \geq 0)$ is Markov by the independence of increments of

$(L(t), t \geq 0)$.

Now we discuss the stationarity of the process $(\mathbf{X}(t), t \geq 0)$, where we restrict our concern to weak stationarity. A multivariate stochastic process $(\mathbf{V}(t), t \geq 0)$ is called *weakly stationary* if its mean is constant and the covariance matrix $Cov(\mathbf{V}(t), \mathbf{V}(s))$ depends on time only through $t - s$, for any $t, s \in [0, \infty)$. With $\sigma(t)$ being constant, the necessary and sufficient condition for stationarity of $\{\mathbf{X}(t)\}$ is, see Brockwell [2009], Proposition 1,

**Theorem 3.2.** *If* $\{\mathbf{X}(0)\}$ *is independent of* $\{L(t), t \geq 0\}$ *and* $\mathbb{E}(L(1)^2) < \infty$*, then* $\{\mathbf{X}(t)\}$ *is weakly stationary if and only if the eigenvalues of the matrix* $\mathbf{A}$ *all have strictly negative real parts and* $\mathbf{X}(0)$ *has the mean and covariance matrix of* $\int_0^\infty e^{Au}\mathbf{e}_p\sigma dL(u)$.

Theorem (3.2) holds in the case of $\sigma = \sigma(t)$, if we further impose that $\sigma(t)$ is bounded.

## 3.2 A Link to Discrete-time Autoregressive Process

In this section we focus on how to relate a CAR process to a discrete-time AR process. The links between the CAR process and its discrete counterpart will be used for analyzing the discrete sampling of a CAR process at uniformly spaced times.

For $\mathbf{X}(t)$ in equation (3.7) with $X_q(t)$ being its $q$th coordinate, we have the stochastic differential equation of $X_q(t)$ for each $q$ separately, i.e.,

$$dX_q(t) = X_{q+1}dt, \quad \text{for } q = 1, 2, \ldots, p-1,$$

$$dX_p(t) = -\sum_{q=1}^{p} \alpha_{p-q+1} X_q dt + \sigma(t) dL(t), \quad \text{for } q = p.$$

Using an Euler approximation of the stochastic differential equations with time step of length one, i.e. $dt = 1$, we have

$$X_q(t+r) - X_q(t+r-1) = X_{q+1}(t+r-1), \tag{3.11}$$

for $q = 1, 2, \ldots, p-1$ and $r \geq 1$. And

$$X_p(t+1) - X_p(t) = -\sum_{q=1}^{p} \alpha_{p-q+1} X_q(t) + \sigma(t)e(t). \tag{3.12}$$

where $e(t) := L(t+1) - L(t)$. By calculating iteratively using equations (3.11) and (3.12), we have the following result.

**Theorem 3.3.** *The following recursive expression for $X_1$ holds,*

$$\begin{aligned}&\sum_{k=0}^{p}(-1)^k c_k^p X_1(t+p-k)\\ =&-\sum_{q=1}^{p} a_{p-q+1}\sum_{k=0}^{q-1}(-1)^k c_k^{q-1} X_1(t+q-1-k)+\sigma(t)e(t),\end{aligned} \tag{3.13}$$

*with coefficients $c_k^q$ defined recursively as*

$$c_k^q = c_{k-1}^{q-1} + c_k^{q-1}, \quad k = 1, 2, \ldots, p-1, q \geq 2,$$

*and $c_0^q = c_q^q = 1$ for $q = 0, 1, \ldots, p$.*

*Proof.* see Benth et al. [2008] , page 284. □

Reorganizing (3.13), we can obtain an expression of $X_1(t)$ in the form of AR($p$) process

$$X_1(t) = \sum_{i=1}^{p} b_i X_1(t-i) + \sigma(t)e(t). \tag{3.14}$$

As we have seen, the discrete sampling of the $X_1(t)$ at time $t = 0, 1, \ldots$ produces an AR process. Later we will fit the observations of $X_1(t)$ to an AR process and identify the coefficients $a_i$, $i = 1, 2, \ldots, p$ for the CAR process from coefficients $b_i, i = 1, 2, \ldots, p$ of the associating AR process.

Moreover, we will see that CAR(3) and CAR(4) processes fit our data sufficiently well. Therefore we include the examples for CAR(3) and CAR(4) here to introduce how the parameter $a_i$ is identified.

**Example 3** (CAR(3))**.** *Let $p = 3$, we have*

$$\begin{aligned}X_1(t+3) = &(3-a_1)X_1(t+2) + (2a_1 - a_2 - 3)X_1(t+1)\\ &+ (-a_1 + a_2 - a_3 + 1)X_1(t) + \sigma(t)e(t).\end{aligned}$$

**Example 4** (CAR(4))**.** *Let $p = 4$, we have*

$$\begin{aligned}X_1(t+4) = &(-a_1+4)X_1(t+3) + (3a_1 - a_2 - 6)X_1(t+2)\\ &+ (-3a_1 + 2a_2 - a_3 + 4)X_1(t+1) + (a_1 - a_2 + a_3 - a_4 - 1)X_1(t) + \sigma(t)e(t).\end{aligned}$$

## 3.3 Model for Daily Average Temperature

In this section, we specify the model for the daily average temperature (DAT). The model includes a seasonal mean and a continuous-time autoregressive (CAR) process with lag $p$ and seasonal variance. It is first suggested by Benth et al. [2007] under the context of weather derivatives, with Brownian motion being the driving noise process. Later the family of generalized hyperbolic Lévy processes are considered as the driving noise in order to capture negative skewness and heavy tails in the residuals.

Here we try to build up a model hopefully to reproduce all the characteristic features of the temperature data. Before that, we list several assumptions that are made and verified on the DAT in most of the literatures. First of all, strong seasonality exists and the DAT is moving around the seasonal mean. A change in the mean of DAT over time is expected due to Global warming and urbanization effect. Autoregressive changes in DAT are assumed. And we further assume that the temperature volatility are time-varying.

We model the DAT $T(t)$ at time $t \geq 0$ as

$$T(t) = \Lambda(t) + Y(t), \tag{3.15}$$

where $\Lambda(t)$ is a deterministic function denoting the seasonal mean; $Y(t)$ is a stochastic process modelling the random fluctuations around the mean. $Y(t) = T(t) - \Lambda(t)$ is called the *deseasonalized* temperatures. The motivation behind this is based on the observation that the DATs always show a strong seasonal cycle with relatively small fluctuations around it. Therefore, most studies model the seasonal mean first and the fluctuations separately. Here we model seasonal mean with deterministic functions $\Lambda(t)$ and hope that the deseasonalized temperature $Y(t)$ is a stationary process with mean 0. The stochastic variability of temperatures is then entirely generated by the process $Y(t)$, as posed in Jewson and Brix [2005].

The seasonal mean function $\Lambda(t)$ should include *trend* and *seasonal effect*. By seasonal effect, we mean the seasonal oscillation in temperature which is cyclic every year. By trend, we mean possible increase in temperature due to global warming and urbanization. Here we use Fourier truncated series (FTS) to model the seasonal mean, that is

$$\Lambda(t) = c_0 + c_1 t + \sum_{r=1}^{R} \left( c_{2r} \cos\left(2\pi r \frac{t}{365}\right) + c_{2r+1} \sin\left(2\pi r \frac{t}{365}\right) \right). \tag{3.16}$$

Parameter $c_0$ is the average of temperature over a year. $c_1$ indicates the global warming or urban effect. Under this framework, we do not include 29th of February in the leap

years. The period of oscillation is 365 days, therefore we have the phase $2\pi t/365$. A simplification is to assume that $R = 1$ as in Alaton et al. [2002]. But commonly the choice for $R$ is relaxed to provide more fine-tuning.

Campbell and Diebold [2005] argues two advantages of using FTS. One is that it produces a smooth pattern which is appropriate to describe the gradual change in the DAT. The other is that its formulated in a simple way which ensures numerical stability in estimation.

The deseasonalized temperatures $Y(t)$ is modelled by a CAR process with seasonal variance, that is

$$Y(t) = \mathbf{e}_1^\top \mathbf{X}(t) = X_1(t), \tag{3.17}$$

$$d\mathbf{X}(t) = \mathbf{A}\mathbf{X}(t)dt + \mathbf{e}_p \sigma(t) dL(t), \tag{3.18}$$

where $\mathbf{e}_k$ the $k$-th unit vector in $\mathbb{R}^p$ for $k = 1, \ldots, p$; $\mathbf{e}^\top$ is the transpose of the vector $\mathbf{e}$; $X_1(t)$ is the first element of the $\mathbf{X}(t)$ ; $L = L(t)$ is a Lévy process; $\sigma(t)$ is a deterministic function; $\mathbf{A}$ is a $p \times p$-matrix

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 & \ldots & 0 & 0 \\ 0 & 0 & 1 & \ldots & 0 & 0 \\ \vdots & & \ddots & & 1 & \vdots \\ 0 & \ldots & \ldots & 0 & 0 & 1 \\ -a_p & -a_{p-1} & \ldots & & -a_2 & -a_1 \end{bmatrix}, \tag{3.19}$$

with constants $a_k$ for $k = 1, \ldots, p$.

The general CAR process allows a time dependent deterministic function $\sigma(t)$ which models the seasonality in the temperature variance, since a periodic oscillation with period of 365 days is often observed in the residuals of temperature after removing the autoregressive effects. With similar consideration as for the seasonal mean $\Lambda(t)$, we use FTS to model the volatility $\sigma(t)$ by

$$\sigma^2(t) = d_1 + \sum_{m=1}^{M} \left( d_{2m} \cos\left(2\pi m \frac{t}{365}\right) + d_{2m+1} \sin\left(2\pi m \frac{t}{365}\right) \right). \tag{3.20}$$

Parameters $d_1$ is the mean of variance of the DAT; $d_2, d_3, \ldots$ characterize the amplitude and shape of the oscillation; and $2\pi t/365$ is the phase indicating a period of 365 days.

$\mathbf{X}(t)$ has explicit solution as in equation (3.9). As given in equation (3.10), $\mathbf{X}(t)$ also satisfies

$$\mathbf{X}(s) = e^{\mathbf{A}(s-t)}\mathbf{X}(t) + \int_t^s e^{\mathbf{A}(s-u)} \mathbf{e}_p \sigma(u) dL(u),$$

therefore $T(t)$ satisfies

$$T(s) = \Lambda(s) + \mathbf{e}_1^\top e^{\mathbf{A}(s-t)}\mathbf{X}(t) + \int_t^s \mathbf{e}_1^\top e^{\mathbf{A}(s-u)}\mathbf{e}_p \sigma(u) dL(u), \tag{3.21}$$

for $s \geq t \geq 0$.

If the eigenvalues of the matrix $\mathbf{A}$ have negative real parts, $e^{\mathbf{A}t} \to 0$ as $t \to 0$. Further if the expectation of the Wiener-Lévy integral in equation (3.10) is constructed to be zero, $\mathbf{X}(t)$ has a mean equal to zero vector, see Swishchuk and Cui [2013]. Then the process $T(t)$ is mean reverting to $\Lambda(t)$. In other words, the temperature mainly depends on the seasonal mean and tends to move back towards the seasonal mean.

The deseasonalized temperature estimated daily is a discrete sampled process ($Y(t) = X_1(t), t = 0, 1, \dots$). As discussed before in section 3.2, it has a form of an AR process

$$X_1(t) = \sum_{i=1}^{p} b_i X_1(t-i) + \epsilon(t), \tag{3.22}$$

and

$$\epsilon(t) = \sigma(t)e(t), \tag{3.23}$$

where $\sigma(t)$ is the volatility; $b_i$, $i = 1, \dots, p$ are constants determined by $a_i$, $i = 1, \dots, p$ in the matrix $\mathbf{A}$; $e(t) = L(t+1) - L(t)$, $t = 1, 2, \dots$ are Lévy increments independently and identically distributed as $L(1)$.

We assume the driving noise process is a Lévy process $L(t)$. Moreover, we consider four cases: a Brownian motion, a generalized hyperbolic (GHYP) Lévy process, a hyperbolic (HYP) Lévy process, and a normal inverse Gaussian (NIG) process. Respectively, $e(t)$ has standard normal distribution, GHYP distribution, HYP distribution and NIG distribution. The motivation behind is that the estimated $e(t)$ always possesses negative skewness and heavy tails, which can not be captured by Brownian motion. The family of GHYP distributions provides a manageable class of distributions which allows these properties.

# Chapter 4

# Temperature Derivatives

Given on our temperature model, here we explore the prices for several temperature futures and options under the arbitrage theory in this chapter.

In section 1, we define the arbitrage-free prices of temperature futures and options. Then in section 2, the Esscher transform is introduced to specify a class of equivalent martingale measures (EMMs). The following section presents the explicit pricing formula for cumulative average temperature (CAT) futures. Section 4 introduces explicit pricing formula for CAT options in the case of Brownian motion, and the Fourier approach to estimate the price in the case of general Lévy processes. In section 5, we discuss pricing for cooling degree days (CDD) and heating degree days (HDD) futures. The issues of the market price of risk (MPR) are discussed in section 6.

## 4.1 Arbitrage-Free Pricing of Derivatives

Here we define the arbitrage-free prices of temperature futures and options.

In the CME market we have temperature futures written on CAT, HDD and CDD indices introduced in section 1.1. We rewrite the temperature indices in a continuous way for the convenience for pricing

$$CAT(\tau_1, \tau_2) = \int_{\tau_1}^{\tau_2} T(t)dt, \tag{4.1}$$

$$CDD(\tau_1, \tau_2) = \int_{\tau_1}^{\tau_2} (T(t) - c)^+ dt, \tag{4.2}$$

$$HDD(\tau_1, \tau_2) = \int_{\tau_1}^{\tau_2} (c - T(t))^+ dt, \tag{4.3}$$

where $(x)^+ = max(x, 0)$, $[\tau_1, \tau_2]$ is the measurement period, $c$ is the reference temperature

for CDD and HDD indices, and $T(t)$ is the daily average temperature (DAT) on time $t$. $T(t)$ is modelled as defined in section 3.3 by a Lévy driven CAR process with seasonal volatility, where $T(t)$ satisfies

$$T(s) = \Lambda(s) + \mathbf{e}_1^\top e^{\mathbf{A}(s-t)}\mathbf{X}(t) + \int_t^s \mathbf{e}_1^\top e^{\mathbf{A}(s-u)}\mathbf{e}_p\sigma(u)dL(u),$$

for $s \geq t \geq 0$. Here $\Lambda(s)$ and $\sigma(s)$ are deterministic functions given by equation (3.16) and equation (3.20).

Temperature market is a classical example of incomplete markets since the underlying temperature indices for temperature futures are not tradable. Moreover, we consider Lévy processes in our model and this introduces a higher level of incompleteness in our market model. Here we rely on the arbitrage theory to study the price of temperature derivatives. Under the no-arbitrage assumption, the price processes of all tradable assets in the temperature market should be arbitrage-free. An equivalent martingale measure (EMM) $Q$ exists, which is a probability measure equivalent to $P$ such that the discounted price processes of all tradable assets are martingales with respect to $Q$. Because the market is incomplete, there are infinitely many EMMs. For the market of temperature futures, the risk-free bond with interest rate $r$ is the only tradable asset. Using the bond price as a numéraire, the discounted bond price is 1 and it is a trivial martingale with respect to all equivalent measures. Thus, all equivalent measures $Q$ are EMMs.

The temperature futures are traded in the market and their price processes should be arbitrage-free. The buyer of the futures enters the contract at time $t$ and will receive a random amount $Y$ in return of futures price $F(t, \tau_1, \tau_2)$. The total payoff at the end of the measurement period $\tau_2$ is $Y - F(t, \tau_1, \tau_2)$. Because the cost to enter the futures contract is zero, by the risk-neutral valuation formula in Theorem (2.19), the futures price $F(t, \tau_1, \tau_2)$ satisfies

$$0 = e^{-r(\tau_2 - t)}\mathbb{E}_Q[Y - F(t, \tau_1, \tau_2)|\mathcal{F}_t], \tag{4.4}$$

for each time $t$, where $Q$ is an EMM measure. The futures price is then given by

$$F(t, \tau_1, \tau_2) = \mathbb{E}_Q[Y|\mathcal{F}_t]. \tag{4.5}$$

Therefore, the prices of the CAT, CDD and HDD futures at time $t \leq \tau_1 \leq \tau_2$ are, respectively,

$$F_{CAT}(t, \tau_1, \tau_2) = \mathbb{E}_Q\left[\int_{\tau 1}^{\tau 2} T(u)du \mid \mathcal{F}_t\right], \tag{4.6}$$

$$F_{CDD}(t,\tau_1,\tau_2) = \mathbb{E}_Q\left[\int_{\tau 1}^{\tau 2}(T(u)-c)^+du \mid \mathcal{F}_t\right], \tag{4.7}$$

$$F_{HDD}(t,\tau_1,\tau_2) = \mathbb{E}_Q\left[\int_{\tau 1}^{\tau 2}(c-T(u))^+du \mid \mathcal{F}_t\right]. \tag{4.8}$$

Moreover, we have the following CDD-HDD parity.

**Proposition 4.1.** *The prices of CAT, HDD and CDD futures have the following relationship*

$$F_{CDD}(t,\tau_1,\tau_2) - F_{HDD}(t,\tau_1,\tau_2) = F_{CAT}(t,\tau_1,\tau_2) - c(\tau_1-\tau_2). \tag{4.9}$$

*Proof.* The result follows by $(T(t)-c)^+ - (c-T(t))^+ = T(t)-c$, □

The options in the CME market are written on temperature futures. As an example, we consider a call option written on $F(\tau,\tau_1,\tau_2)$ at time $\tau \le \tau_1 \le \tau_2$ with strike $K$. The buyer enters the call option at time $t \le \tau$ and receives a payoff of $(F(\tau,\tau_1,\tau_2)-K)^+$ at time $\tau$. The price process $\{F(t,\tau_1,\tau_2), t \in [0,\tau_{max}]\}$ defined by equation (4.4) is a martingale under the measure $Q$ in its definition. Therefore, once we have pinned down a pricing measure $Q$ for the futures price, the option price is given by

$$C(t,\tau,\tau_1,\tau_2) = e^{-r(\tau-t)}\mathbb{E}_Q\left[(F(\tau,\tau_1,\tau_2)-K)^+ \mid \mathcal{F}_t\right]. \tag{4.10}$$

## 4.2 Equivalent Measures by Esscher Transform

In such an incomplete market, the prices defined above are not unique and depend on $Q$. As discussed before, every equivalent measure is an EMM measure in our model. The way to single out one EMM $Q$ is to restrict to a parameterized class of equivalent measures and fit the theoretically derived prices to the observed prices. If a Brownian motion is the driving noise process, the Girsanov transform is applied for the change of measure. As Lévy process is included is included in our model, we apply the Esscher transform which is a generalization of the Girsanov transform.

Before applying the Esscher transform, we first impose the following integrability conditions on the Lévy measure. The following condition is to ensure that the underlying asset process has finite moments up to some orders, see Swishchuk and Cui [2013], page 91, Condition 1.

**Condition 4.2.** *There exists a constant $k > 0$ such that the Lévy measure $\ell$ satisfies the integrability condition*

$$\int_{|z|\ge 1} \mid z \mid^k \ell(dz) < \infty$$

*almost surely.*

The price is the expectation of the discounted process under a risk-neutral measure $Q$ defined by Esscher transform, which requires an exponential integrability, see Benth and Šaltytė-Benth [2004], page 186, Condition (L).

**Condition 4.3.** *There exists a constant $k > 0$ such that the Lévy measure $\ell$ satisfies the integrability condition*

$$\int_{|z|\geq 1} z^{kz}\ell(dz) < \infty.$$

With the exponential integrability, we have the following Lemma

**Lemma 4.4.** *Let function $f : [0,T] \mapsto \mathbb{R}$ be a bounded and measurable function and Condition 4.3 holds for $k := sup_{s\in[0,t]}|f(s)|$, then*

$$\mathbb{E}\left[\exp\left(\int_0^t f(s)dL(s)\right)\right] = \exp\left(\int_0^t \psi(f(s))ds\right),$$

*where $\psi(u) = \eta(-iu)$, $\eta$ is the Lévy symbol.*

*Proof.* see Benth and Šaltytė-Benth [2004], page 186, Lemma 4.1. □

Recall that the characteristic function of $L(1)$ is $\phi_{L(1)}(u) = \mathbb{E}[e^{iuL(1)}] = e^{\eta(u)}$, with Lévy symbol $\eta(u)$. Then $\mathbb{E}[e^{uL(1)}] = e^{\eta(-iu)} = e^{\psi(u)}$, i.e. $\psi(u)$ is the logarithm of the moment generating function of $L(1)$.

Now we introduce the class of EMMs via Esscher transform. Let $\tau_{max}$ denote a fixed time horizon including the trading time for all relevant temperature derivatives. Let $\theta : [0,\tau_{max}] \mapsto \mathbb{R}$ is a measurable and bounded function. Define for $t \in [0,\tau_{max}]$ the stochastic process

$$Z^\theta(t) = \exp\left(\int_0^t \theta(s)dL(s) - \int_0^t \psi(\theta(s))ds\right), \tag{4.11}$$

where $\psi(u)$ is the logarithm of the moment generating function of $L(1)$. The process $Z^\theta(t)$ is well defined for $t \in [0,\tau_{max}]$ if Condition 4.3 holds for $k := sup_{s\in[0,t]}|\theta(s)|$. The probability measure $Q^\theta$ is defined such that $Z^\theta(t)$ is the density process of the Radon-Nikodym derivative $\frac{dQ^\theta}{dP}$, i.e.

$$\frac{dQ^\theta}{dP}|\mathcal{F}_t = Z^\theta(t). \tag{4.12}$$

Or equivalently,

$$Q^\theta(A) = \mathbb{E}\left[1_A Z^\theta(\tau_{max})\right], \tag{4.13}$$

where $1_A$ is the indicator function. Now we have a class of equivalent martingale measures $Q^\theta$ which is parametrized by $\theta$. If the Lévy process is a Brownian motion, the Esscher transform corresponds to the Girsanov transform. The expectation under $Q^\theta$ is denoted by $\mathbb{E}_Q[\cdot]$. It is worth mentioning that Esscher Transform defined in this way is dependent on the driving Lévy process.

As argued in Benth and Benth [2012], we restrict to the Esscher transform in order to obtain a flexible class of EMMs tractable for pricing and at the same time can be used for estimation of the market risk premium. The Esscher transform introduces a class of parametrized EMMs. We will see later that the prices of some derivatives can be represented explicitly in terms of the cumulant function of the underlying Lévy process and the parameter $\theta$. Therefore, the parameters $\theta$ can be estimated from observed market price and interpreted as the market price of risk.

## 4.3 Futures on Cumulative Average Temperature

The price of the CAT futures, as given in equation (4.6), is

$$F_{CAT}(t, \tau_1, \tau_2) = \mathbb{E}_Q \left[ \int_{\tau 1}^{\tau 2} T(u) du \mid \mathcal{F}_t \right]. \tag{4.14}$$

We will see later that under the Lévy driven CAR model the price of CAT futures has an closed form formula. Before we derive the pricing formula of CAT futures, we need an Lemma by Benth and Šaltytė-Benth [2005].

**Lemma 4.5.** *Given a measurable and bounded function $f(t)$, the expectation of Lévy integral under risk neutral measure $Q$ over $[\tau_1, \tau_2]$ is*

$$\mathbb{E}_Q \left[ \int_{\tau_1}^{\tau_2} f(t) dL(t) \mid \mathcal{F}_{\tau_1} \right] = \int_{\tau_1}^{\tau_2} f(t) \psi'(\theta(t)) dt,$$

*where $\psi(u)$ is the logarithm of the moment generating function of $L(1)$, $\psi'(u)$ is the derivative of $\psi(u)$ with respect to $u$.*

*Proof.* Using the Bayes Formula, the independent increment property of Lévy process and

Lemma 4.4, we have

$$\begin{aligned}
&\mathbb{E}_Q\left[\int_{\tau_1}^{\tau_2} f(t)dL(t) \mid \mathcal{F}_{\tau_1}\right] \\
=&\mathbb{E}\left[\int_{\tau_1}^{\tau_2} f(t)dL(t)\frac{Z^\theta(\tau_2)}{Z^\theta(\tau_1)} \mid \mathcal{F}_{\tau_1}\right] \\
=&\mathbb{E}\left[\int_{\tau_1}^{\tau_2} f(t)dL(t) \times \exp\left(\int_{\tau_1}^{\tau_2}\theta(s)dL(s) - \int_{\tau_1}^{\tau_2}\psi(\theta(s))ds\right) \mid \mathcal{F}_{\tau_1}\right] \\
=&\exp\left(-\int_{\tau_1}^{\tau_2}\psi(\theta(s))ds\right)\frac{d}{d\lambda}\mathbb{E}\left[\int_{\tau_1}^{\tau_2}\lambda f(s) + \theta(u)dL(s)\right] \mid_{\lambda=0} \\
=&\exp\left(-\int_{\tau_1}^{\tau_2}\psi(\theta(s))ds\right)\frac{d}{d\lambda}\exp\left(\int_{\tau_1}^{\tau_2}\psi(\lambda f(s) + \theta(u))ds\right) \mid_{\lambda=0} \\
=&\int_{\tau_1}^{\tau_2} f(t)\psi'(\theta(t))dt
\end{aligned}$$

□

With the above Lemma, we present the pricing formula in following Theorem, see Swishchuk and Cui [2013], page 94, Theorem 6.

**Theorem 4.6.** *The price of CAT futures for* $0 \le t \le \tau_1 \le \tau_2$ *is*

$$\begin{aligned}
F_{CAT}(t,\tau_1,\tau_2) = &\int_{\tau_1}^{\tau_2}\Lambda(u)du + \mathbf{a}(t,\tau_1,\tau_2)\mathbf{X}(t) + \int_t^{\tau_1}\sigma(u)\psi'(\theta(u))\mathbf{a}(t,\tau_1,\tau_2)\mathbf{e}_p du \\
&+ \int_{\tau_1}^{\tau_2}\sigma(u)\psi'(\theta(u))\mathbf{e}_1^\top\mathbf{A}^{-1}\left(e^{\mathbf{A}(\tau_2-u)} - I_p\right)\mathbf{e}_p du, \qquad (4.15)
\end{aligned}$$

*where* $\mathbf{a}(t,\tau_1,\tau_2) = \mathbf{e}_1^\top\mathbf{A}^{-1}\left(e^{\mathbf{A}(\tau_2-t)} - e^{\mathbf{A}(\tau_1-t)}\right)$*;* $I_p$ *is a* $p \times p$ *identity matrix;* $\psi(u)$ *is the logarithm of the moment generating function of* $L(1)$*,* $\psi'(u)$ *is the derivative of* $\psi(u)$ *with respect to* $u$*.*

*Proof.* In order to derive

$$F_{CAT}(t,\tau_1,\tau_2) = \mathbb{E}_Q\left[\int_{\tau 1}^{\tau 2} T(u)du \mid \mathcal{F}_t\right]$$

given in equation (4.6), we first derive the dynamics of $\int_{\tau 1}^{\tau 2} T(u)du$.

Integrate equation (3.7) on $[\tau_1,\tau_2]$, we have

$$\mathbf{X}(\tau_2) - \mathbf{X}(\tau_1) = \mathbf{A}.\int_{\tau_1}^{\tau_2}\mathbf{X}(t)dt + \mathbf{e}_p\int_{\tau_1}^{\tau_2}\sigma(t)dL(t).$$

Using equation (3.10) for $\mathbf{X}(\tau_1)$ and $\mathbf{X}(\tau_2)$, we have

$$\begin{aligned}\int_{\tau_1}^{\tau_2}\mathbf{X}(u)du =&\mathbf{A}^{-1}\left(\mathbf{X}(\tau_2)-\mathbf{X}(\tau_1)-\mathbf{e}_p\int_{\tau_1}^{\tau_2}\sigma(u)dL(u)\right)\\ =&\mathbf{A}^{-1}\left(\left(e^{\mathbf{A}(\tau_2-t)}-e^{\mathbf{A}(\tau_1-t)}\right)\mathbf{X}(t)+\int_t^{\tau_2}e^{\mathbf{A}(\tau_2-u)}\mathbf{e}_p\sigma(u)dL(u)\right.\\ &\left.-\int_t^{\tau_1}e^{\mathbf{A}(\tau_1-u)}\mathbf{e}_p\sigma(u)dL(u)-\mathbf{e}_p\int_{\tau_1}^{\tau_2}\sigma(u)dL(u)\right).\end{aligned}$$

Hence

$$\begin{aligned}&\int_{\tau_1}^{\tau_2}T(u)du\\ =&\int_{\tau_1}^{\tau_2}\left(\Lambda(u)+\mathbf{e}_1^\top\mathbf{X}(u)\right)du\\ =&\int_{\tau_1}^{\tau_2}\Lambda(u)du+\mathbf{e}_1^\top\int_{\tau_1}^{\tau_2}\mathbf{X}(u)du\\ =&\int_{\tau_1}^{\tau_2}\Lambda(u)du+\mathbf{e}_1^\top\mathbf{A}^{-1}\left(\left(e^{\mathbf{A}(\tau_2-t)}-e^{\mathbf{A}(\tau_1-t)}\right)\mathbf{X}(t)+\int_t^{\tau_2}e^{\mathbf{A}(\tau_2-u)}\mathbf{e}_p\sigma(u)dL(u)\right.\\ &\left.-\int_t^{\tau_1}e^{\mathbf{A}(\tau_1-u)}\mathbf{e}_p\sigma(u)dL(u)-\mathbf{e}_p\int_{\tau_1}^{\tau_2}\sigma(u)dL(u)\right)\end{aligned}$$

We calculate the conditional expectation of each term separately. From Lemma 4.5, we have for example

$$\mathbb{E}_Q\left[\int_t^{\tau_2}\mathbf{e}_1^\top\mathbf{A}^{-1}e^{\mathbf{A}(\tau_2-u)}\mathbf{e}_p\sigma(u)dL(u)\mid\mathcal{F}_t\right]=\int_t^{\tau_2}\sigma(u)\psi'(\theta(u))\mathbf{e}_1^\top\mathbf{A}^{-1}e^{\mathbf{A}(\tau_2-u)}\mathbf{e}_p du$$

Summing up the expectations of each term leads to the pricing formula. □

In the case when the Lévy process $L$ is a standard Brownian motion $B$, the logarithm of the moment generating function of $B(1)$ is $\psi_{BM}(\theta)=\frac{1}{2}\theta^2$ for each $\theta\in\mathbb{R}$. $\psi'_{BM}(\theta)=\theta$, then we have

**Example 5.** *If the driving noise process $L$ is a Brownian motion, the price of CAT futures*

$$\begin{aligned}F_{CAT}(t,\tau_1,\tau_2)=&\int_{\tau_1}^{\tau_2}\Lambda(u)du+\mathbf{a}(t,\tau_1,\tau_2)\mathbf{X}(t)+\int_t^{\tau_1}\theta(u)\sigma(u)\mathbf{a}(t,\tau_1,\tau_2)\mathbf{e}_p du\\ &+\int_{\tau_1}^{\tau_2}\theta(u)\sigma(u)\mathbf{e}_1^\top\mathbf{A}^{-1}\left(e^{\mathbf{A}(\tau_2-u)}-I_p\right)\mathbf{e}_p du,\end{aligned}\tag{4.16}$$

*where* $\mathbf{a}(t,\tau_1,\tau_2)=\mathbf{e}_1^\top\mathbf{A}^{-1}\left(e^{\mathbf{A}(\tau_2-t)}-e^{\mathbf{A}(\tau_1-t)}\right)$*; $I_p$ is a $p\times p$ identity matrix.*

## 4.4 Options on Cumulative Average Temperature Futures

Here we discuss the pricing of options written on CAT futures. The price of a call option on future price $F_{CAT}(\tau, \tau_1, \tau_2)$ at time $t \leq \tau$ with strike $K$ is

$$C_{CAT}(t, \tau, \tau_1, \tau_2) = e^{-r(\tau-t)}\mathbb{E}_Q\left[(F_{CAT}(\tau, \tau_1, \tau_2) - K)^+ \mid \mathcal{F}_t\right]. \tag{4.17}$$

Thus, we require the distribution of $F_{CAT}(\tau, \tau_1, \tau_2)$ under risk-neutral measure $Q$. From equation (4.15), the distribution of $F_{CAT}(\tau, \tau_1, \tau_2)$ depends on $\mathbf{X}(\tau)$, where

$$\mathbf{X}(\tau) = e^{\mathbf{A}(\tau-t)}\mathbf{X}(t) + \int_t^\tau e^{\mathbf{A}(\tau-u)}\mathbf{e}_p\sigma(u)dL(u), \tag{4.18}$$

for $\tau > t$.

In the special case when Lévy process $L$ is a Brownian motion $B$, we can derive an explicit pricing formula for CAT options based on the insight that $\mathbf{X}(\tau)$ is normally distributed under $Q$. An explicit pricing formula for a call option written on $F_{CAT}(t, \tau_1, \tau_2)$ is derived by Benth et al. [2007], as in following Theorem.

**Theorem 4.7.** *If $L$ is a Brownian motion $B$, the price of a call option at time $t$ written on a CAT futures with strike price $K$ and exercise date $\tau$, where the measure period of the futures is $[\tau_1, \tau_2]$ with $t \leq \tau \leq \tau_1 < \tau_2$, is*

$$\begin{aligned} C_{CAT}(t, , \tau, \tau_1, \tau_2) =& e^{-r(\tau-t)} \times \{(F_{CAT}(t, \tau_1, \tau_2) - K)\Phi(w(t, \tau, \tau_1, \tau_2)) \\ &+ \int_t^\tau \Sigma^2_{CAT}(s, \tau_1, \tau_2)ds\Phi'(w(t, \tau, \tau_1, \tau_2))\} \end{aligned} \tag{4.19}$$

*where*

$$w(t, \tau, \tau_1, \tau_2) = \frac{F_{CAT}(t, \tau_1, \tau_2) - K}{\sqrt{\int_t^\tau \Sigma^2_{CAT}(s, \tau_1, \tau_2)ds}}$$

*and $\Phi$ is the cumulative standard normal distribution function.*

*Proof.* The Esscher transform coincides with Girsanov transform in this case, so that

$$B^\theta(t) := B(t) - \int_0^t \theta(u)du$$

is a Brownian motion under under $Q = Q^\theta$ for $t \in [0, \tau_{max}]$. The dynamics of $\mathbf{X}(t)$ under $Q$ becomes

$$d\mathbf{X}(t) = (\mathbf{A}\mathbf{X}(t) + \mathbf{e}_p\sigma(t)\theta(t))\,dt + \mathbf{e}_p\sigma(t)dB^\theta(t). \tag{4.20}$$

The CAT futures dynamics under the risk neutral measure $Q^\theta$ is then

$$dF_{CAT}(t,\tau_1,\tau_2) = \Sigma_{CAT}(t,\tau_1,\tau_2)dB^\theta(t) \tag{4.21}$$

where

$$\Sigma_{CAT}(t,\tau_1,\tau_2) = \sigma(t)\mathbf{e}_1^\top \mathbf{A}^{-1}\left(e^{\mathbf{A}(\tau_2-t)} - e^{\mathbf{A}(\tau_1-t)}\right)\mathbf{e}_p. \tag{4.22}$$

This is true using the fact that $F_{CAT}(t,\tau_1,\tau_2)$ is a $Q^\theta$-martingale, see Benth et al. [2007], Proposition 2. Following equation (4.21), we have

$$F_{CAT}(\tau,\tau_1,\tau_2) = F_{CAT}(t,\tau_1,\tau_2) + \int_t^\tau \Sigma_{CAT}(s,\tau_1,\tau_2)dB^\theta(s),$$

for $\tau \geq t$. Thus, $F_{CAT}(\tau,\tau_1,\tau_2)$ conditioned on $F_{CAT}(t,\tau_1,\tau_2)$ is normally distributed under $Q$. The result follows the properties of normal distribution, see Benth et al. [2007], Proposition 3. □

From equation (4.19), we notice that the option price no longer depends on $\theta$ if $L$ is a Brownian motion particularly. Therefore, after we single out a measure $Q^\theta$ for CAT futures, the market consisting of CAT futures and options is complete. Then we can replicate the payoff of an option with the underlying futures using a delta hedging strategy, see Benth et al. [2007].

For general Lévy processes, it is not straightforward to derive explicit pricing formula for equation (4.17), because the explicit distribution of $\mathbf{X}(\tau)$ is not known under $Q$, except for some very special cases. The difficulty lies in finding an explicit density function for the Wiener integral of the form

$$\int_t^u e^{\mathbf{A}(u-s)}\mathbf{e}_p\sigma(s)dL(s), \quad \text{for } u > t. \tag{4.23}$$

As an example, we consider $L$ being a hyperbolic Lévy process with $L(1)$ following hyperbolic distribution. Hyperbolic distributions are not closed under convolution, as posed in Bellini [2005]. Therefore, we can not find an exact expression for density of $L(t)$ for $t \neq 1$. As a consequence, explicit density for a process including $L(t)$ for $t \neq 1$ is not feasible.

The common approach to evaluate option price (4.17) is to use numerical integration or Monte Carlo simulation. Here we discuss the method using a Fourier approach. The motivation behind is that the characteristic function of the form (4.23) admits explicit expression, as presented in the following Lemma.

**Lemma 4.8.** *For a constant vector* $\mathbf{c} \in \mathbb{R}^p$*, the characteristic function of the random*

*variable*

$$G(u) := \int_t^u \mathbf{c}^\top e^{\mathbf{A}(u-s)}\mathbf{e}_p\sigma(s)dL(s)$$

*under the risk neutral measure $Q$ is given by*

$$\phi_{G(u)}(\lambda) = \mathbb{E}_Q[exp(i\lambda G(u))|\mathcal{F}_t] = \exp\left\{\Theta(\lambda)\right\}, \tag{4.24}$$

*with*

$$\Theta(\lambda) = \int_t^u \psi\left(i\lambda\mathbf{c}^\top e^{\mathbf{A}(u-s)}\mathbf{e}_p\sigma(s) + \theta(s)\right)ds - \int_t^u \psi\left(\theta(s)\right)ds, \tag{4.25}$$

*and $\psi(u)$ being the logarithm of the moment generating function of $L(1)$.*

*Proof.* Using the Bayes Formula, the independent increment property of Lévy process and Lemma 4.4, we have

$$\begin{aligned}
&\mathbb{E}_Q\left[\exp\left\{\mathbf{c}^\top i\lambda\int_t^u e^{\mathbf{A}(u-s)}\mathbf{e}_p\sigma(s)dL(s)\right\} \mid \mathcal{F}_t\right] \\
=&\mathbb{E}\left[\exp\left\{i\lambda\int_t^u \mathbf{c}^\top e^{\mathbf{A}(u-s)}\mathbf{e}_p\sigma(s)dL(s)\right\}\frac{Z^\theta(u)}{Z^\theta(t)} \mid \mathcal{F}_t\right] \\
=&\mathbb{E}\left[\exp\left\{i\lambda\int_t^u \mathbf{c}^\top e^{\mathbf{A}(u-s)}\mathbf{e}_p\sigma(s)dL(s)\right.\right. \\
&\left.\left.+\int_t^u \theta(s)dL(s) - \int_t^u \psi(\theta(s))ds\right\} \mid \mathcal{F}_t\right] \\
=&\exp\left\{-\int_t^u \psi(\theta(s))ds\right\} \\
&\cdot\mathbb{E}\left[\exp\left\{i\lambda\int_t^u \mathbf{c}^\top e^{\mathbf{A}(u-s)}\mathbf{e}_p\sigma(s)dL(s) + \int_t^u \theta(s)dL(s)\right\}\right] \\
=&\exp\left\{\int_t^u \psi\left(i\lambda\mathbf{c}^\top e^{\mathbf{A}(t-u)}\mathbf{e}_p\sigma(s) + \theta(s)\right)ds - \int_t^u \psi(\theta(s))ds\right\}
\end{aligned}$$

□

We now present some details on how the Fourier approach is conducted to evaluate the price. We follow the framework developed by Swishchuk and Cui [2013] on pricing HDD and CDD futures.

First we review the mathematical insights and the general construction of this method. The following Theorem links the density function of a random variable to its characteristic function.

**Theorem 4.9.** *Let $F(x)$ denote the cumulative density function of a random variable $X$, the corresponding probability density function $f(x)$ is integrable in Lebesgue sense, i.e.*

*$f(x) \in \mathcal{L}$, The characteristic function of $X$ is defined as $\phi(t) = \int_{-\infty}^{+\infty} e^{itx} f(x)dx \in \mathcal{L}$. Then*

$$\begin{aligned} f(x) &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-itx} \phi(t) dt \\ &= \frac{1}{\pi} \mathrm{Re} \left[ \int_{0}^{+\infty} e^{-itx} \phi(t) dt \right], \end{aligned} \tag{4.26}$$

*where* $\mathrm{Re}[x]$ *denotes the real part of a complex number* $x$*.*

*Proof.* The result follows from Fourier Inversion Theorem, see Hewitt and Ross [1963], page 409. □

We follow the construction introduced in Chourdakis [2004], section 1, section 2. The integral in equation (4.26) can be approximated by

$$f(x) \approx \frac{1}{\pi} \mathrm{Re} \left[ \sum_{k=0}^{N-1} \delta_k e^{-it_k x} \phi(t_k) \Delta t \right], \tag{4.27}$$

where $\delta_k, k = 0, 1, \ldots, N-1$ are the integration weights implementing some integration rule. We choose the rule such that $\delta_k = \frac{1}{2}$ for $k = 0$ and $\delta_k = 1$ otherwise. The points $t_k$ are chosen to be equidistant with grid spacing $\Delta t$, then $t_k = k\Delta t$. The value of $\Delta t$ should be sufficiently small to approximate the integral well enough, while the value of $N\Delta t$ should be large enough to assume that the characteristic function is equal to zero for $t > \bar{t} = N\Delta t$.

Application of discrete Fourier transform (DTF) will result in a set of numbers approximating equation (4.26) on the points $\{x_k, k = 0, \ldots, N-1\}$. The points $\{x_k\}$ are set to be equidistant with grid spacing $\Delta x$. Then the values $x_k$ is of the form

$$x_k = -b + k\Delta x, \quad k = 0, \ldots, N-1, \tag{4.28}$$

where $b$ is a parameter to control the return range. For example, if $b = \frac{\Delta x N}{2}$, $\{x_k\}$ evenly spread around $x = 0$. By contrasting equation (4.27) with DTF transform, we have

$$\Delta x \Delta t = \frac{2\pi}{N} \Rightarrow \Delta x = \frac{2\pi}{N\Delta t} \tag{4.29}$$

so that

$$\begin{aligned} f(x_j) &\approx \mathrm{Re}\left[\frac{1}{\pi}\sum_{k=0}^{N-1}\delta_k e^{-it_k x_j}\phi(k\Delta t)\Delta t\right] \\ &= \mathrm{Re}\left[\sum_{k=0}^{N-1}\frac{1}{\pi}\delta_k e^{-i(k\Delta t)(-b+j\Delta x)}\phi(k\Delta t)\Delta t\right] \\ &= \mathrm{Re}\left[\sum_{k=0}^{N-1}\frac{1}{\pi}\delta_k e^{ikb\Delta t}e^{-ik\frac{2\pi}{N}j}\phi(k\Delta t)\Delta t\right] \\ &= \mathrm{Re}\left[\sum_{k=0}^{N-1} f_k e^{-ik\frac{2\pi}{N}j}\right] \end{aligned} \tag{4.30}$$

for $j = 0, \ldots, N-1$, where

$$f_k := \frac{1}{\pi}\delta_k e^{i\Delta tkb}\phi(\Delta tk)\Delta t, \quad k = 0, 1, \ldots, N-1. \tag{4.31}$$

From equation (4.30), the sequence $\{f(x_j), j = 0, \ldots, N-1\}$ is the real parts of the DTF of $\{f_k, k = 0, \ldots, N-1\}$, which can be computed fastly by the Fast Fourier Transform.

Back to the price of CAT option, insert the futures price $F_{CAT}(\tau, \tau_1, \tau_2)$ given by equation (4.15) and the dynamics of $\mathbf{X}(\tau)$ by equation (4.18) into the option price (4.17), we have

$$\begin{aligned} &C_{CAT}(t, \tau, \tau_1, \tau_2) \\ =&e^{-r(\tau-t)}\mathbb{E}_Q\left[(F_{CAT}(\tau, \tau_1, \tau_2) - K)^+ \mid \mathcal{F}_t\right] \\ =&e^{-r(\tau-t)}\mathbb{E}_Q\left[\left(k(t, \tau, \tau_1, \tau_2) + \int_t^\tau \mathbf{a}(\tau, \tau_1, \tau_2)e^{\mathbf{A}(\tau-u)}\mathbf{e}_p\sigma(u)dL(u)\right)^+ \mid \mathcal{F}_t\right] \\ =&e^{-r(\tau-t)}\int_{-k(t,\tau,\tau_1,\tau_2)}^{+\infty}(k(t, \tau, \tau_1, \tau_2) + x)f(x)dx \end{aligned} \tag{4.32}$$

where

$$\begin{aligned} k(t, \tau, \tau_1, \tau_2) := &\int_{\tau_1}^{\tau_2}\Lambda(u)du + \int_\tau^{\tau_1}\sigma(u)\psi'(\theta(u))\mathbf{a}(\tau, \tau_1, \tau_2)\mathbf{e}_p du \\ &+ \int_{\tau_1}^{\tau_2}\sigma(u)\psi'(\theta(u))\mathbf{e}_1^\top\mathbf{A}^{-1}\left(e^{\mathbf{A}(\tau_2-u)} - I_p\right)\mathbf{e}_p du \\ &+ \mathbf{a}(\tau, \tau_1, \tau_2)e^{\mathbf{A}(\tau-t)}\mathbf{X}(t) - K, \end{aligned} \tag{4.33}$$

and $f(x)$ is the density of

$$\int_t^\tau \mathbf{a}(\tau, \tau_1, \tau_2) e^{\mathbf{A}(\tau-u)} \mathbf{e}_p \sigma(u) dL(u). \tag{4.34}$$

Here we outline the algorithm to estimate the call option on CAT futures. Let $\phi(u)$ be the characteristic function of the term (4.34) which is known from Lemma 4.8. After choosing suitable values $\Delta x$, $N$ and the $b$, we construct the sequence $\{f_k, k = 0, \ldots, N-1\}$ as given in equation (4.31). Then we perform DTF to obtain $\{f(x_j), j = 0, \ldots, N-1\}$, which is the approximations of $f(x)$ on points $\{x_j, j = 0, \ldots, N-1\}$. The CAT option price is given numerically by

$$C_{CAT}(t, \tau, \tau_1, \tau_2) \approx e^{-r(\tau-t)} \sum_{x_j > -k(t,\tau,\tau_1,\tau_2)} (k(t, \tau, \tau_1, \tau_2) + x_j) f(x_j) \Delta x, \tag{4.35}$$

where $k(t, \tau, \tau_1, \tau_2)$ is given by (4.33).

## 4.5 Futures on Cooling and Heating Degree Day

As given in equations (4.7) and (4.8), the prices of CDD and HDD futures are

$$\begin{aligned} F_{CDD}(t, \tau_1, \tau_2) &= \mathbb{E}_Q \left[ \int_{\tau 1}^{\tau 2} (T(u) - c)^+ du \mid \mathcal{F}_t \right] \\ &= \int_{\tau 1}^{\tau 2} \mathbb{E}_Q \left[ (T(u) - c)^+ \mid \mathcal{F}_t \right] du \\ &= \int_{\tau 1}^{\tau 2} \int_c^{+\infty} (x - c) f_{T(u)}(x) dx du, \end{aligned} \tag{4.36}$$

and

$$\begin{aligned} F_{HDD}(t, \tau_1, \tau_2) &= \mathbb{E}_Q \left[ \int_{\tau 1}^{\tau 2} (c - T(u))^+ du \mid \mathcal{F}_t \right] \\ &= \int_{\tau 1}^{\tau 2} \int_{-\infty}^{c} (c - x) f_{T(u)}(x) dx du, \end{aligned} \tag{4.37}$$

where $f_{T(u)}(x)$ is the density function of $T(u)$ under $Q$. Thus, in order to calculate the price of a CDD/HDD future, we have to know the distribution of $T(u), u \in [\tau_1, \tau_2]$ under $Q$.

In the special case when $L$ is Brownian motion, the distribution of $T(u), u \in [\tau_1, \tau_2]$ is normal under $Q$. The appealing properties of normal distributions allow explicit pricing

formulas for CDD and HDD futures. The formulas, derived by Benth et al. [2007], are given in the following Theorem.

**Theorem 4.10.** *If $L$ is a Brownian motion, the CDD future price*

$$F_{CDD}(t,\tau_1,\tau_2) = \int_{\tau_1}^{\tau_2} \nu(t,s)\Psi\left[\frac{m(t,s)-c}{\nu(t,s)}\right] ds, \tag{4.38}$$

*and the HDD future price*

$$F_{HDD}(t,\tau_1,\tau_2) = \int_{\tau_1}^{\tau_2} \nu(t,s)\Psi\left[\frac{c-m(t,s))}{\nu(t,s)}\right] ds, \tag{4.39}$$

*where*

$$m(t,u) = \Lambda(u) + \mathbf{e}_1^\top e^{\mathbf{A}(u-t)}\mathbf{X}(t) + \int_t^u \theta(s)\sigma(s)\mathbf{e}_1^\top e^{\mathbf{A}(u-s)}\mathbf{e}_p ds, \tag{4.40}$$

$$\nu^2(t,u) = \int_t^u \sigma^2(s)\left(\mathbf{e}_1^\top e^{\mathbf{A}(u-s)}\mathbf{e}_p\right)^2 ds, \tag{4.41}$$

*$\Phi$ is the cumulative standard normal distribution function and $\Psi(x) = x\Phi(x) + \Phi'(x)$.*

*Proof.* We prove for CDD futures here. First we calculate the inner integral in equation (4.36). Under Girsanov transform, $\mathbf{X}(t)$ has solution

$$\mathbf{X}(t) = e^{\mathbf{A}(t-s)}\mathbf{X}(s) + \int_s^t e^{\mathbf{A}(t-u)}\mathbf{e}_p\sigma(u)\theta(u)du + \int_s^t e^{\mathbf{A}(t-u)}\mathbf{e}_p\sigma(u)dB^\theta(u), \tag{4.42}$$

for $t \geq s \geq 0$, where $B^\theta$ is a Brownian motion under $Q$. Thus, $T(u) = \Lambda(u) + \mathbf{e}_1^\top\mathbf{X}(u)$ is normally distributed under $Q$ with mean $m(t,u) = \mathbb{E}_Q[T(u)|\mathcal{F}_t]$ given by (4.40) and variance $\nu^2(t,u) = Var_Q[T(u)|\mathcal{F}_t]$ given by (4.41)

Using properties for normal distribution, $S := \frac{T(u)-m}{\nu} \sim N(0,1)$, and its density function $f_S(s) = \Phi'(s) = \exp\{\frac{1}{2}s^2\}$, where $\Phi$ is the cumulative standard normal distribution function. Let $s = \frac{x-m}{\nu}$ and $\tilde{c} := \frac{c-m}{\nu}$, we have

$$\begin{aligned}\int_c^{+\infty}(x-c)f_{T(u)}(x)dx &= \int_{\tilde{c}}^{+\infty}(\nu s + m - c)f_S(s)ds \\ &= \nu\int_{\tilde{c}}^{+\infty} s f_S(s)ds + (m-c)\int_{\tilde{c}}^{+\infty} f_S(s)ds\end{aligned}$$

where

$$\nu\int_{\tilde{c}}^{+\infty} s f_S(s)ds = \nu\int_{\tilde{c}}^{+\infty} s\exp\{\frac{1}{2}s^2\}ds = -\nu\int_{\tilde{c}}^{+\infty} df_S(s) = \nu\Phi'(-\tilde{c}),$$

$$(m-c)\int_{\tilde{c}}^{+\infty} f_S(s)ds = (m-c)\Phi(-\tilde{c}) = -\tilde{c}\nu\Phi(-\tilde{c})$$

Sum up two terms and define $\Psi(x) := x\Phi(x) + \Phi'(x)$, we have the final expression. □

Finding a simple closed form solution for price of HDD/CDD future is difficult for general Lévy processes with non-Gaussian increments. The argument is similar to pricing CAT options. The difficulty lies in finding an exact expression for the density of $T(s)$ under $Q$, typically with the existence of the term

$$\int_t^s \mathbf{e}_1^\top e^{\mathbf{A}(t-u)}\mathbf{e}_p\sigma(u)dL(u) \tag{4.43}$$

in $T(s)$ from equation (3.21). However, the characteristic function of (4.43) is explicitly known from Lemma (4.8). As a consequence, the characteristic function of $T(s)$ under $Q$ is known, as shown in the following Proposition.

**Proposition 4.11.** *The characteristic function of $T(s)$ for $s \geq t$ conditioned on $\mathcal{F}_t$ under the risk neutral measure $Q$ is given by*

$$\phi_{T(s)}(\lambda) = \mathbb{E}_Q\left[\exp\{i\lambda T(s)\} \mid \mathcal{F}_t\right] = \exp\{\Theta_{T(s)}(\lambda)\},$$

*where*

$$\begin{aligned}\Theta_{T(s)}(\lambda) =& i\lambda\Lambda(s) + i\lambda\mathbf{e}_1^\top e^{\mathbf{A}(s-t)}\mathbf{X}(t) - \int_t^s \psi\left(\theta(u)\right)du \\ &+ \int_t^s \psi\left(i\lambda\mathbf{e}_1^\top e^{\mathbf{A}(s-u)}\mathbf{e}_p\sigma(u) + \theta(u)\right)du,\end{aligned}$$

*with $\psi(u)$ being the logarithm of the moment generating function of $L(1)$.*

*Proof.* Simple calculation with Lemma 4.8. □

Therefore, we can apply the Fourier technique based on characteristic function of $T(t)$ to estimate the price. Here we outline a method to estimate the price using Fourier transform concerning the term (4.43). This approach is introduced in Swishchuk and Cui [2013].

For price of a CDD future, we have

$$\begin{aligned}
&F_{CDD}(t,\tau_1,\tau_2)\\
&=\int_{\tau 1}^{\tau 2}\mathbb{E}_Q\left[(T(u)-c)^+\mid\mathcal{F}_t\right]du\\
&=\int_{\tau 1}^{\tau 2}\mathbb{E}_Q\left[\left(\Lambda(u)+\mathbf{e}_1^\top e^{\mathbf{A}(u-t)}\mathbf{X}(t)+\underbrace{\int_t^u\mathbf{e}_1^\top e^{\mathbf{A}(u-s)}\mathbf{e}_p\sigma(s)dL(s)}_{=:G(u)}-c\right)^+\mid\mathcal{F}_t\right]du\\
&=\int_{\tau 1}^{\tau 2}\int_{d(u)}^{+\infty}(x-d(u))f_{G(u)}(x)dxdu,
\end{aligned}\tag{4.44}$$

where $d(u)=c-\Lambda(u)-\mathbf{e}_1^\top e^{\mathbf{A}(u-t)}\mathbf{X}(t)$ and $f_{G(u)}(x)$ is the density of $G(u)$. From Lemma 4.8, the characteristic function of $G(u)$ conditioned on $\mathcal{F}_t$ under the risk neutral measure $Q$ is given by

$$\phi_{G(u)}(\lambda)=\exp\left\{\Theta_{G(u)}(\lambda)\right\}\tag{4.45}$$

where

$$\Theta_{G(u)}(\lambda)=\int_t^u\psi\left(i\lambda\mathbf{e}_1^\top e^{\mathbf{A}(u-s)}\mathbf{e}_p\sigma(s)+\theta(s)\right)ds-\int_t^u\psi\left(\theta(s)\right)ds,\tag{4.46}$$

with $\psi(u)$ being the logarithm of moment generating function of $L(1)$. We construct $\{x_k,k=1,\ldots,N-1\}$ in the same way as introduced for pricing CAT option and estimate $f_{G(u)}(x_k)$ with $\phi_{G(u)}$ for $u\in[\tau_1,\tau_2]$. Then the CDD and HDD futures prices are given numerically as follows. The price of HDD future is

$$F_{CDD}(t,\tau_1,\tau_2)\approx\sum_{u=\tau 1}^{\tau 2}\sum_{x_k\geq d(u)}\left[x_k+\left(\Lambda(u)+\mathbf{e}_1^\top e^{\mathbf{A}(u-t)}\mathbf{X}(t)-c\right)\right]f_{G(u)}(x_k)\Delta x\Delta u,\tag{4.47}$$

and the price of HDD future is

$$F_{HDD}(t,\tau_1,\tau_2)\approx\sum_{u=\tau 1}^{\tau 2}\sum_{x_k\leq d(u)}\left[\left(c-\Lambda(u)-\mathbf{e}_1^\top e^{\mathbf{A}(u-t)}\mathbf{X}(t)\right)-x_k\right]f_{G(u)}(x_k)\Delta x\Delta u,\tag{4.48}$$

where $d(u)=c-\Lambda(u)-\mathbf{e}_1^\top e^{\mathbf{A}(u-t)}\mathbf{X}(t)$. This corrects the numerical representation given by Swishchuk and Cui [2013], page 95, Theorem 5.

Because the CDD and HDD futures prices are given in a numerical way, we do not include discussion on the pricing of CDD and HDD options here. In the special case of Brownian motion, one can refer to Benth et al. [2007]. More generally, Monte Carlo simulations can be used.

## 4.6 The Market Price of Risk

Here we introduce the concept of the market price of risk, and how it can be further explored to study the prices behavior in the market.

The temperature derivatives market is a classical incomplete market. In incomplete markets, there are infinite many EMMs. Before we have introduced a parametric class of measures $Q^\theta$ by Esscher transform, where parameter $\theta = \theta(t)$ is square integrable functions in our assumption. The price of the derivatives depends on the selected EMM, and hence on the parameter $\theta(t)$. $\theta(t)$ is interpreted as the market price of risk (MPR). The choice of $\theta(t)$ reflects the preference towards risk of the market participants, as posed in Bellini [2005].

Recent studies have found that the MPR is different than zero. In order to study the significance of the MPR, we can calibrate the MPR from market data. The calibration process is in fact the inverse problem to the pricing process. The estimated value of $\theta$ gives the theoretical prices that are closest to the market prices according to some minimizing criterion.

Assume we have different futures contracts $i = 1, \ldots, I$ with measurement period $[\tau_1, \tau_2]$. For a single trading date $t$, or the so called close date, we observe market prices $F^*(t, \tau_1^i, \tau_2^i)$ for contracts $i = 1, \ldots, I$. We restrict to the case where the MPR is constant for each contract per trading date, i.e. the MPR is constant over $[t, \tau_2^i]$ for each contract $i = 1, \ldots, I$. The aim to minimize the objective function given as the squared distance between the observed market price $F^*(t, \tau_1^i, \tau_2^i)$ and theoretical price $F(t, \tau_1^i, \tau_2^i)$ estimated under the pricing framework, i.e. we solve the problem

$$\min \left\{ F(t, \tau_1^i, \tau_2^i) - F^*(t, \tau_1^i, \tau_2^i) \right\}^2 .$$

For example, since we have an closed form pricing formula for CAT futures under the Lévy driven CAR model, we have that

$$\begin{aligned} \theta_{CAT}^i(t) = \arg\min_{\theta^i(t)} \Bigg( & F_{CAT}^*(t, \tau_1, \tau_2) - \int_{\tau_1}^{\tau_2} \Lambda(u) du - \mathbf{a}(t, \tau_1, \tau_2) \mathbf{X}(t) \\ & - \psi'(\theta^i(t)) \left\{ \int_t^{\tau_1} \sigma(u) \mathbf{a}(t, \tau_1, \tau_2) \mathbf{e}_p du \right. \\ & \left. + \int_{\tau_1}^{\tau_2} \sigma(u) \mathbf{e}_1^\top \mathbf{A}^{-1} \left( e^{\mathbf{A}(\tau_2 - u)} - I_p \right) \mathbf{e}_p du \right\} \Bigg)^2 \end{aligned} \tag{4.49}$$

Hence we obtain the MPR $\theta^i(t)$ which is varying with respect to time $t$ but it is constant on $[t, \tau_2^i]$.

This framework can be generalized with more complex formulations on $\theta(t)$. For example, Härdle and Cabrera [2012] consider formulations of one piece constant, two piece constant, and a general form of series expansion for $\theta$ under the Brownian driven CAR model. We can put similar considerations here for CAT futures, since it also has an explicit expression with cumulant function of the Lévy process. It is worth mentioning that one usually calibrates the MPR for temperature futures and use the implied MPR for further pricing of options.

In fact, the market price of risk is associated with the economic concept of *risk premium*, defined as the difference between the risk neutral price $F(t, \tau_1, \tau_2)$ which is the expected pay-off under $Q$, and the price $F_P(t, \tau_1, \tau_2)$ which the expected payoff under physical measure $P$, i.e.,

$$RP = F(t, \tau_1, \tau_2) - F_P(t, \tau_1, \tau_2). \tag{4.50}$$

For more discussion on the economic interpretation, one can refer to Härdle and Cabrera [2012].

# Chapter 5

# Temperature Data Analysis

In this chapter, we analyze historical temperature data of 6 cites in the United States (US) and 2 cities in Europe. We fit the data to the continuous-time autoregressive (CAR) model and the results are presented here.

In the first section, we give the source of our temperature data and some descriptive statistics of the data. In the second section, we describe the overall fitting procedures. The last few sections are organized in the way to present the results for each step in the procedures separately.

## 5.1 Data Description

We have historical daily temperature data of 2 European cities and 6 US cities, which are Berlin, Stockholm, Atlanta, Chicago, Dallas, New York, Philadelphia and Portland. For US cities, the data set is obtained from the National Climatic data Center (NCDC) under the National Oceanic and Atmospheric Administration (NOAA).[1] Historical data of Berlin is obtained from the Deutscher Wetterdienst (DWD).[2] Data of Stockholm is obtained from Swedish Meteorological and Hydrological Institute (SMHI).[3] These sources offer historical temperature data measured on different stations for each city, and we select the right one according to the contracts offered in Chicago Mercantile Exchange (CME).

Observations of daily average temperature (DAT) is calculated as the mean of the maximum and minimum contained in the temperature data. February 29 is removed in every leap year, so that the length of every year is fixed as 365 days. In this thesis, we use the format $mm/dd/yyyy$ for dates, e.g. 09/01/1961 denotes 1st of September in year

[1] `http://www.ncdc.noaa.gov`
[2] `http://www.dwd.de/`
[3] `http://www.smhi.se/`

1961. The data set periods are 01/01/1961 $\sim$ 12/31/2010 for US cities, 01/01/1963 $\sim$ 12/31/2011 for Berlin and 01/01/1961 $\sim$ 12/31/2012 for Stockholm. The data set period lasts for about 50 years, and together we have approximately 18,000 observations for each city. To agree with the notification of the temperature derivatives, the temperature is measured in degree Fahrenheit for US cities and in degree Celsius for European cities.

To give an overview of the data, Figure 5.1 plots the temperature data between 01/01/2006 and 12/31/2010 for Stockholm. Strong seasonality with period of 1 year is observed, with high temperature in the summer and low temperature in the winter. Figure 5.2 plots the histograms for temperature data for all cities. Bimodal distributions are clearly observed for cities such as Berlin, Chicago and Stockholm. The two peaks corresponds to temperature in summer and in winter.

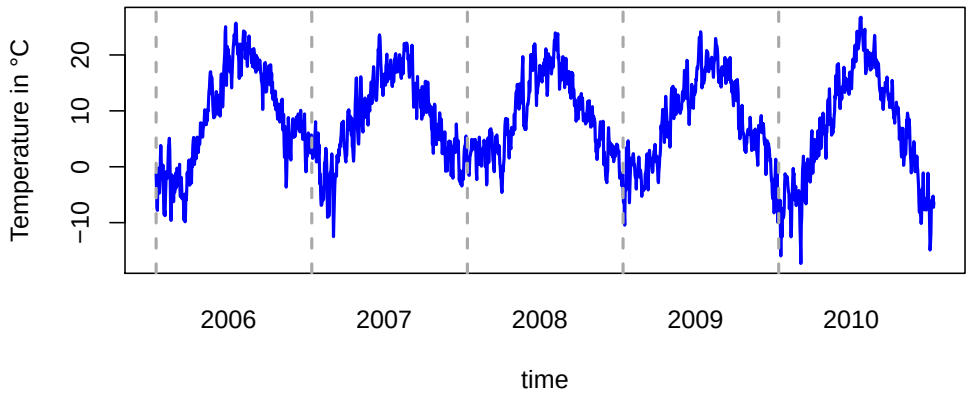



Figure 5.1: Daily average temperatures of Stockholm in period 01/01/2006 $\sim$ 12/31/2010

Table 5.1 presents some descriptive statistics of the temperature data. We see that the statistics varies from city to city and depends on the geological location. The mean temperature ranges from 7.30°C in Stockholm to 65.96°F (18.87°C) in Dallas. The standard deviation is quite large compared to the mean temperature, which indicates that the temperature is volatile. All cities except Portland have negative skewness. Negative excess kurtosis is observed for all cities. This indicates that the distribution of DAT is platykurtic and the mass of distribution is concentrate on the left tail (on the right tail fro Portland). The Jarque-Bera test is performed to test the null hypothesis that the DATs come from a normal distribution. The result is that the null hypothesis is rejected for all cities at significance level of 1%.

Table 5.1: Descriptive statistics of the daily average temperature

| | Mean | Std. dev | Max | Median | Min | Skewness | Kurtosis | JB stat |
|---|---|---|---|---|---|---|---|---|
| Atlanta | 61.92 | 15.05 | 92.03 | 63.95 | 5.00 | -0.43 | -0.74 | 974.26 |
| Chicago | 49.54 | 20.05 | 92.48 | 50.99 | -18.04 | -0.33 | -0.71 | 712.11 |
| Dallas | 65.96 | 16.28 | 96.98 | 67.55 | 8.51 | -0.40 | -0.72 | 887.78 |
| Newyork | 55.14 | 17.35 | 94.01 | 55.49 | 2.57 | -0.19 | -0.91 | 731.10 |
| Philadelphia | 55.18 | 17.72 | 91.94 | 55.49 | 0.50 | -0.21 | -0.95 | 815.82 |
| Portland | 53.98 | 11.55 | 89.96 | 53.06 | 11.03 | 0.02 | -0.56 | 238.77 |
| Berlin | 9.69 | 7.89 | 30.55 | 9.85 | -18.25 | -0.17 | -0.63 | 385.33 |
| Stockholm | 7.30 | 8.23 | 28.70 | 7.00 | -23.60 | -0.11 | -0.68 | 405.52 |

*St. dev* denotes the standard deviation. *JB stat* denotes the Jarque-Bera statistic. The temperature for the first 6 US cities are measured in °F, while that for the last 2 European cities are measured in °C. The critical values for JB test are 5.99 at 5% significance level, and 9.21 at 1% significance level.

## 5.2 Fitting Procedures

Recall from section 3.3 that we have modelled the DAT with a Lévy-driven CAR model. The $T(t)$ at time $t \geq 0$ is

$$T(t) = \Lambda(t) + Y(t),$$

where $\Lambda(t)$ is the seasonal mean given by

$$\Lambda(t) = c_0 + c_1 t + \sum_{r=1}^{R} \left( c_{2r} \cos\left(2\pi r \frac{t}{365}\right) + c_{2r+1} \sin\left(2\pi r \frac{t}{365}\right) \right).$$

and the deseasonalized temperature $Y(t)$ is a general CAR process given by

$$Y(t) = \mathbf{e}_1^\top \mathbf{X}(t),$$

$$d\mathbf{X}(t) = \mathbf{A}\mathbf{X}(t)dt + \mathbf{e}_p \sigma(t) dL(t),$$

where $L = L(t)$ is a Lévy process; $\mathbf{A}$ is a $p \times p$-matrix

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & 1 & \dots & 0 & 0 \\ \vdots & & \ddots & & 1 & \vdots \\ 0 & \dots & \dots & 0 & 0 & 1 \\ -a_p & -a_{p-1} & \dots & & -a_2 & -a_1 \end{bmatrix},$$

with positive constants $a_k$ for $k = 1, \ldots, p$. The discretely sampled deseasonalized temperature $Y(t), t = 1, 2, \ldots$ has the form of an autoregressive process

$$X_1(t) = \sum_{i=1}^{p} b_i X_1(t-i) + \epsilon(t),$$

and

$$\epsilon(t) = \sigma(t)e(t),$$

where $e(t) = L(t+1) - L(t)$ and $\sigma^2(t) > 0$ is the variance modelled by

$$\sigma^2(t) = d_1 + \sum_{m=1}^{M} \left( d_{2m} \cos\left(2\pi m \frac{t}{365}\right) + d_{2m+1} \sin\left(2\pi m \frac{t}{365}\right) \right).$$

We analyzed the temperature data and estimate the parameters of the model in the following ways.

First we remove the seasonal mean in temperature data. The seasonal mean $\Lambda(t)$ is fitted by least squares minimization. Parameters in the model are chosen by minimizing the BIC criteria. Then we have the estimated value of the seasonal mean, $\hat{\Lambda}(t)$. We remove the seasonal mean from the data and hope that the deseasonalized temperature data $\hat{Y}(t) = T(t) - \hat{\Lambda}(t)$ is a stationary process. The ADF test and KPSS test are performed to check stationarity.

Then the deseasonalized temperatures are fitted to an AR($p$) process. The optimal order $p$ is chosen by minimizing some information criteria. Once we have estimated parameters $\hat{b_j}$, $j = 1, 2, \ldots, p$, for the autoregressive process, we identify $\hat{a}_i$, $i = 1, 2, \ldots, p$. Eigenvalues of the matrix $\hat{\mathbf{A}}$ are calculated and the condition for stationarity of the CAR process is examined.

The seasonal volatility is estimated in the following way. From the previous step, we have the residuals $\hat{\epsilon}(t)$ from the AR($p$) process. In order to estimate volatility $\sigma(t)$, we group $\hat{\epsilon}(t)$, $t = 1, 2, \ldots$, into 365 groups for each calendar day $t_i$, $i = 1, 2, \ldots, 365$, and take the average of each groups of $\hat{\epsilon}(t_i)$ as the observations of $\sigma(t_i)$. The parameters in the FTS function for $\sigma(t)$ is estimated by least square minimization with $M$ chosen by minimizing some information criteria.

We call $\hat{e}(t) = \hat{\epsilon}(t)/\hat{\sigma}(t)$ the *standardized residuals*. In last step we estimate the distribution of the standardized residuals. First we test whether the standardized residuals are normally distributed. Further we test whether the standardized residuals come from generalized hyperbolic (GHYP) distribution, hyperbolic (HYP) distribution and normal inverse Gaussian (NIG) distribution.

This stepwise procedure is applied in Benth et al. [2007]. Admittedly, the estimation of the seasonal variance $\sigma^2(t)$ is "*ad hoc*", as posed in Benth et al. [2007]. However, we will see later that it captures seasonal variance sufficiently. Methods for estimation of the AR process with time-dependent residual variance can be found in, for example, Ruppert [1988] and Hayman [1960].

## 5.3 Seasonal Mean

The seasonal mean is modelled by FTS in form of equation (3.16). The parameters under different number of periodic terms, $R$, are first fitted by least squares minimization. We select the number of the number of $R$, by minimizing BIC criteria. Then we drop the parameters that are not significant at a confidence level of 5%, and again aim at minimizing the BIC value. Figure 5.3 plots the observed DAT together with the fitted seasonal function $\hat{\Lambda}(t)$.

Tables 5.2 and 5.3 present the fitted parameters of each city. $\hat{c}_0$ is the mean temperature in a year, ranging from 6.39°C in Berlin to 64.78°F (18.21 °C) in Dallas. $\hat{c}_1$ is positive and non-zero at 5% significance level for all cities, which indicates an increase in the temperature due to global warming or urbanization effects. Though the estimated value of $\hat{c}_1$ is very small in absolute value, it still has significant influence given the large time span, For example, with $\hat{c}_1$ estimated to be 0.0001 for Berlin, the temperature increases by 1.46°C as the time increases by 40 years. We see that $R$ varies from city to city, ranging from 3 in Chicago (Dallas, Newyork and Philadelphia as well) to 8 in Berlin. As posed in Härdle and Cabrera [2012], an increasing number of the periodic terms provides additional flexibility for fine tuning. However, as the number of parameters in increasing, the complexity in estimation is also increased.

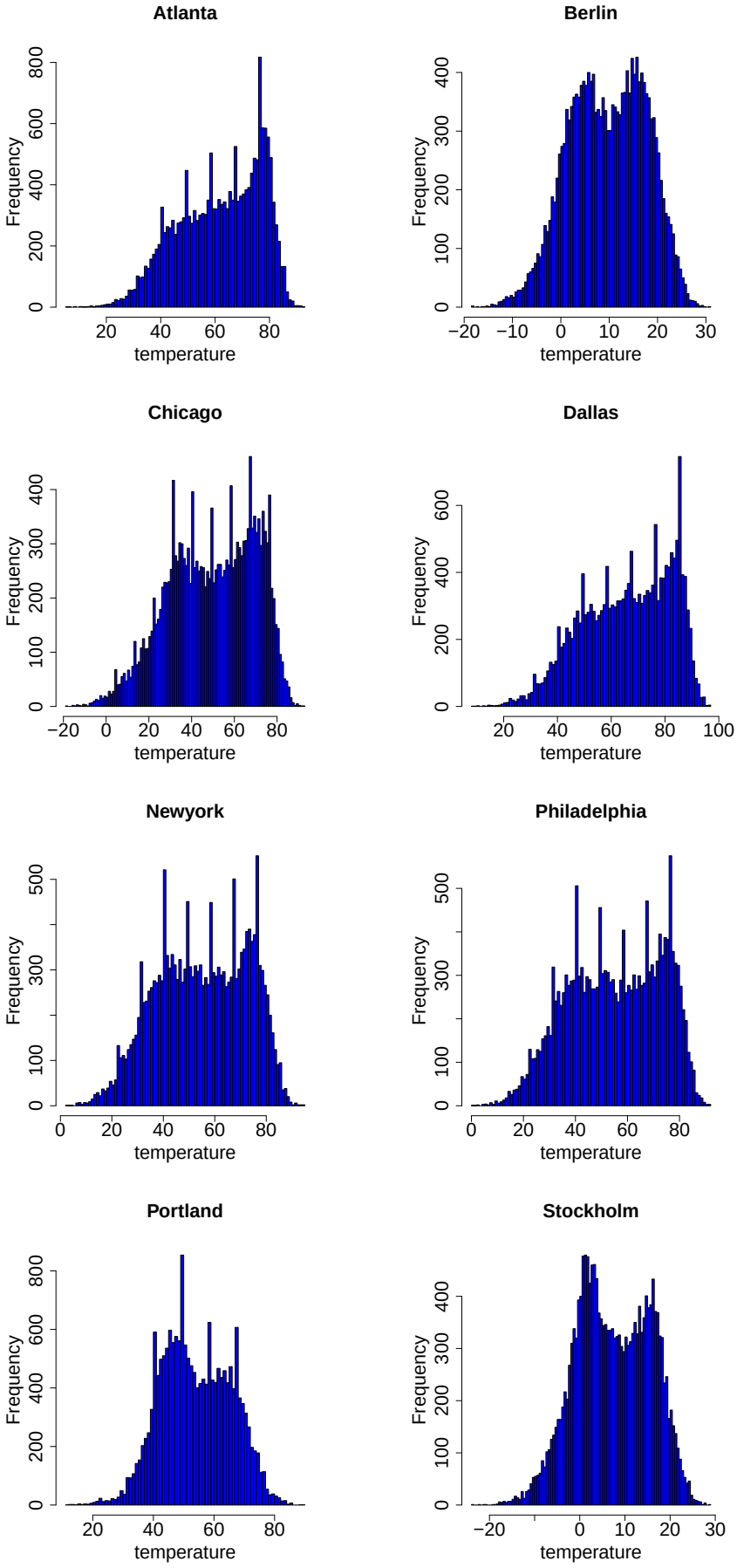



Figure 5.2: Histograms of of daily average temperatures

Table 5.2: Fitted parameters of the seasonal mean (part 1)

| | Atlanta | Chicago | Dallas | Newyork |
|---|---|---|---|---|
| $\hat{c}_0$ (Std. er) | 60.22 (0.11) | 48.51 (0.13) | 64.78 (0.11) | 53.50 (0.10) |
| $\hat{c}_1$ (Std. er) | 0.0001 (0.00001) | 0.0001 (0.00001) | 0.0001 (0.00001) | 0.0001 (0.00001) |
| $\hat{c}_2$ (Std. er) | -17.85(0.07) | -23.75 (0.09) | -19.35 (0.08) | -20.53 (0.07) |
| $\hat{c}_3$ (Std. er) | -5.34 (0.07) | -8.52 (0.09) | -6.06 (0.08) | -8.86 (0.07) |
| $\hat{c}_4$ (Std. er) | -1.32 (0.07) | -1.77 (0.09) | -1.46 (0.08) | -0.35 (0.07) |
| $\hat{c}_5$ (Std. er) | 0.49 (0.07) | - | 1.04 (0.08) | - |
| $\hat{c}_6$ (Std. er) | - | -0.59 (0.09) | -0.40 (0.08) | -0.16 (0.07) |
| $\hat{c}_7$ (Std. er) | -0.20 (0.07) | -0.79 (0.09) | -0.40 (0.08) | -0.72 (0.07) |
| $\hat{c}_8$ (Std. er) | - | - | - | - |
| $\hat{c}_9$ (Std. er) | -0.43 (0.07) | - | - | - |
| $\hat{c}_{10}$ (Std. er) | - | - | - | - |
| $\hat{c}_{12}$ (Std. er) | - | - | - | - |
| $\hat{c}_{13}$ (Std. er) | - | - | - | - |
| $\hat{c}_{14}$ (Std. er) | - | - | - | - |
| $\hat{c}_{15}$ (Std. er) | - | - | - | - |
| $\hat{c}_{16}$ (Std. er) | - | - | - | - |
| $\hat{c}_{17}$ (Std. er) | - | - | - | - |
| $\hat{c}_{18}$ (Std. er) | - | - | - | - |
| $\hat{R}$ | 4 | 3 | 3 | 3 |
| $BIC$ | 123499 | 132090 | 125658 | 123160 |

*Std. er* denotes the standard error. All the coefficients are non-zero at 5% significant level.

Table 5.3: Fitted parameters of the seasonal mean (part 2)

| | Portland | Philadelphia | Berlin | Stockholm |
|---|---|---|---|---|
| $\hat{c}_0$ (Std. er) | 53.00 (0.08) | 53.32 (0.11) | 8.97 (0.05) | 6.39 (0.05) |
| $\hat{c}_1$ (Std. er) | 0.0001 (0.00001) | 0.0004 (0.00001) | 0.0001 (0.00001) | 0.0001 (0.000004) |
| $\hat{c}_2$ (Std. er) | -13.33 (0.05) | -2.12 (0.07) | -9.34 (0.03) | -9.80 (0.03) |
| $\hat{c}_3$ (Std. er) | -5.02 (0.05) | -8.10 (0.07) | -2.91 (0.03) | -3.65 (0.03) |
| $\hat{c}_4$ (Std. er) | -0.74 (0.05) | -0.61 (0.07) | -0.10 (0.03) | 0.75 (0.03) |
| $\hat{c}_5$ (Std. er) | 2.37 (0.05) | -0.79 (0.07) | 0.26 (0.03) | 0.18 (0.03) |
| $\hat{c}_6$ (Std. er) | -0.26 (0.05) | - | 0.16 (0.03) | 0.09 (0.03) |
| $\hat{c}_7$ (Std. er) | - | - | - | -0.12 (0.03) |
| $\hat{c}_8$ (Std. er) | -0.28 (0.05) | - | -0.12 (0.03) | - |
| $\hat{c}_9$ (Std. er) | - | - | 0.14 (0.03) | 0.17 (0.03) |
| $\hat{c}_{10}$ (Std. er) | 0.17 (0.05) | - | 0.15 (0.03) | 0.10 (0.03) |
| $\hat{c}_{12}$ (Std. er) | -0.16 (0.05) | - | - | 0.13 (0.03) |
| $\hat{c}_{13}$ (Std. er) | -0.19 (0.05) | - | -0.19 (0.03) | -0.14 (0.03) |
| $\hat{c}_{14}$ (Std. er) | 0.26 (0.05) | - | - | 0.11 (0.03) |
| $\hat{c}_{15}$ (Std. er) | - | - | -0.15 (0.03) | -0.15 (0.03) |
| $\hat{c}_{16}$ (Std. er) | - | - | - | 0.12 (0.03) |
| $\hat{c}_{17}$ (Std. er) | - | - | - | - |
| $\hat{c}_{18}$ (Std. er) | - | - | -0.20 (0.03) | - |
| $\hat{R}$ | 6 | 3 | 8 | 7 |
| $BIC$ | 112879 | 132090 | 98204 | 101680 |

*Std. er* denotes the standard error. All the coefficients are non-zero at 5% significant level.

After removing the seasonal mean from the temperature data, we perform the ADF test and KPSS test to check whether the deseasonalized temperatures follow a stationary process. From Table 5.4, the Dickey-Fuller statistic is smaller than the critical value at 5% significance level. Hence the null hypothesis of the ADF test that the underlying process is non-stationary is rejected for all cities at 5% significance level. The KPSS statistic is smaller than the critical value. Hence the null hypothesis of the KPSS test that the time series is stationary process cannot be rejected for all cities.

Table 5.4: Results of ADF test and KPSS test for deseasonalized temperature data

| City | Dickey-Fuller | KPSS |
|---|---|---|
| Atlanta | -22.55 | 0.380 |
| Chicago | -22.14 | 0.084 |
| Dallas | -21.65 | 0.271 |
| Newyork | -21.68 | 0.104 |
| Philadelphia | -22.03 | 0.138 |
| Portland | -21.33 | 0.197 |
| Berlin | -21.00 | 0.056 |
| Stockholm | -19.07 | 0.157 |

The critical values for ADF test are -2.86 at 5% significance level, and -3.43 at 1% significance level. The critical values for KPSS test are 0.463 at 5% significance level.

## 5.4 Autoregressive process

After stationarity is verified on the deseasonalized temperature data, we plot the Sample ACFs and Sample PACFs of the deseasonalized temperature data to detect serial dependence. From figures 5.4 and 5.5, we observe that the Sample PACFs display a cutoff at lag 3 while the Sample ACFs decay more slowly for most cities. The plots indicate that AR(3) may be suitable for most cities.

More precisely, we assume the data follows an AR($p$) process and then use an order selection criteria to select the optimal $p$. We use *arima* in R to fit the AR($p$) process, which returns AIC and BIC values. Table 5.5 gives the optimal $p$ with minimum AIC and BIC values, respectively, for each city. The BIC leads to $p = 3$ for all cities except Portland. The AIC leads to different $p$, ranging from 4 for Portland and 12 for Atlanta, Chicago, Dallas and Stockholm. Notice that $p = 4$ is optimal for Portland under both AIC and BIC. Based on the results as well as Sample ACFs and Sample PACFs, we choose $p = 4$ for city Portland, and $p = 3$ for the rest.

Table 5.5: Optimal number of lags $p$ for deseasonalized temperature under AIC and BIC

| | BIC | | AIC | |
|---|---|---|---|---|
| | $p$ | BIC | $p$ | AIC |
| Atlanta | 3 | 107626 | 12 | 107561 |
| Chicago | 3 | 117862 | 12 | 117808 |
| Dallas | 3 | 112619 | 12 | 112571 |
| Newyork | 3 | 111108 | 9 | 111042 |
| Philadelphia | 3 | 111917 | 8 | 111847 |
| Portland | 4 | 98113 | 4 | 98065 |
| Berlin | 3 | 78028 | 8 | 77970 |
| Stockholm | 3 | 80024 | 12 | 79946 |

Table 5.6 lists the fitted coefficients for AR(3) process. As discussed before, we identify the parameters of the matrix $\mathbf{A}$ in the CAR($p$) process from the fitted AR($p$) process. The fitted coefficients for the AR(3) and the parameters calculated for the associated CAR(p) are given in Table 5.6. The real parts of eigenvalues of the fitted matrix are negative for all cities, hence the condition for stationarity is satisfied.

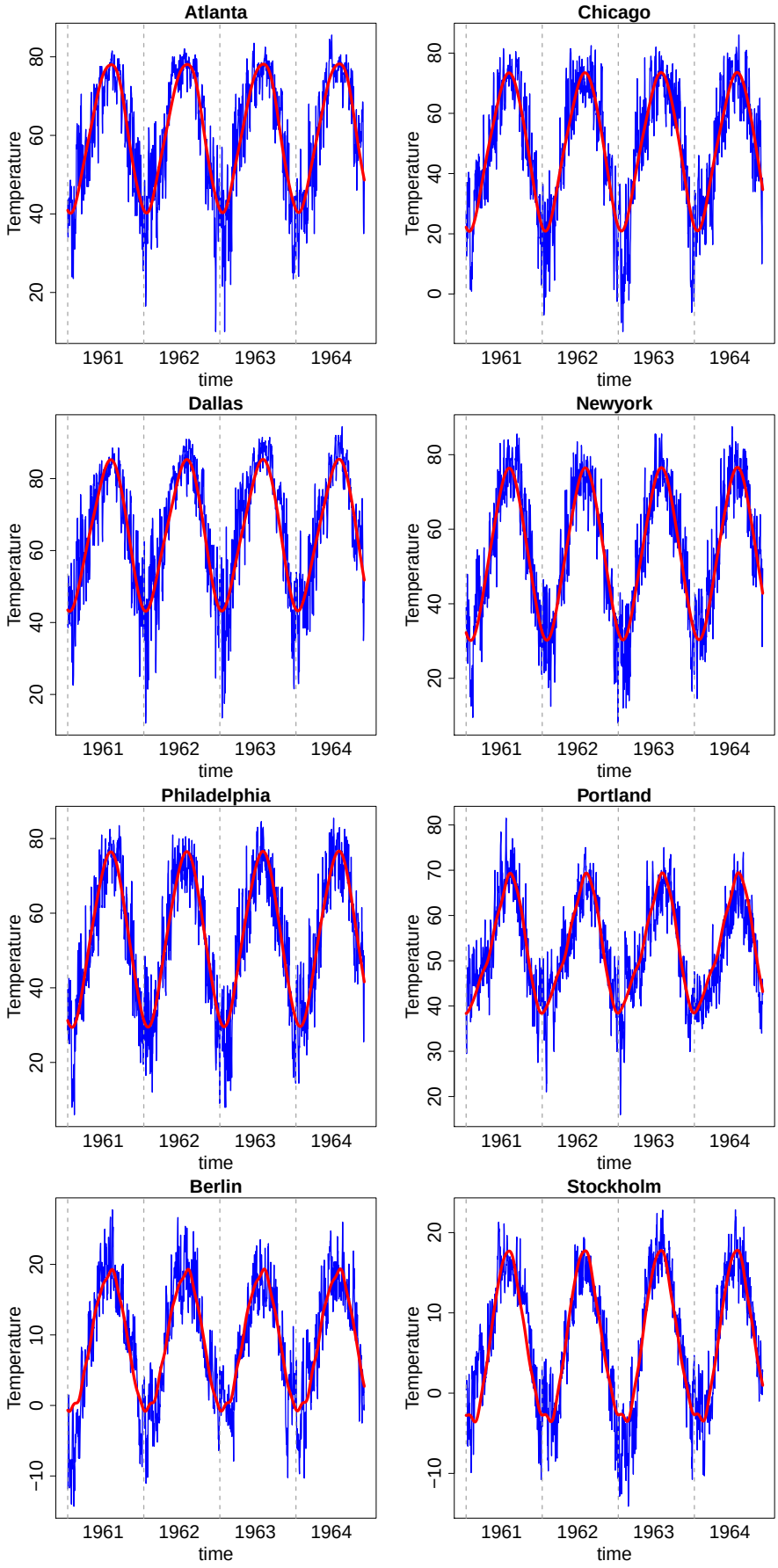



Figure 5.3: The fitted seasonal function over average daily temperatures

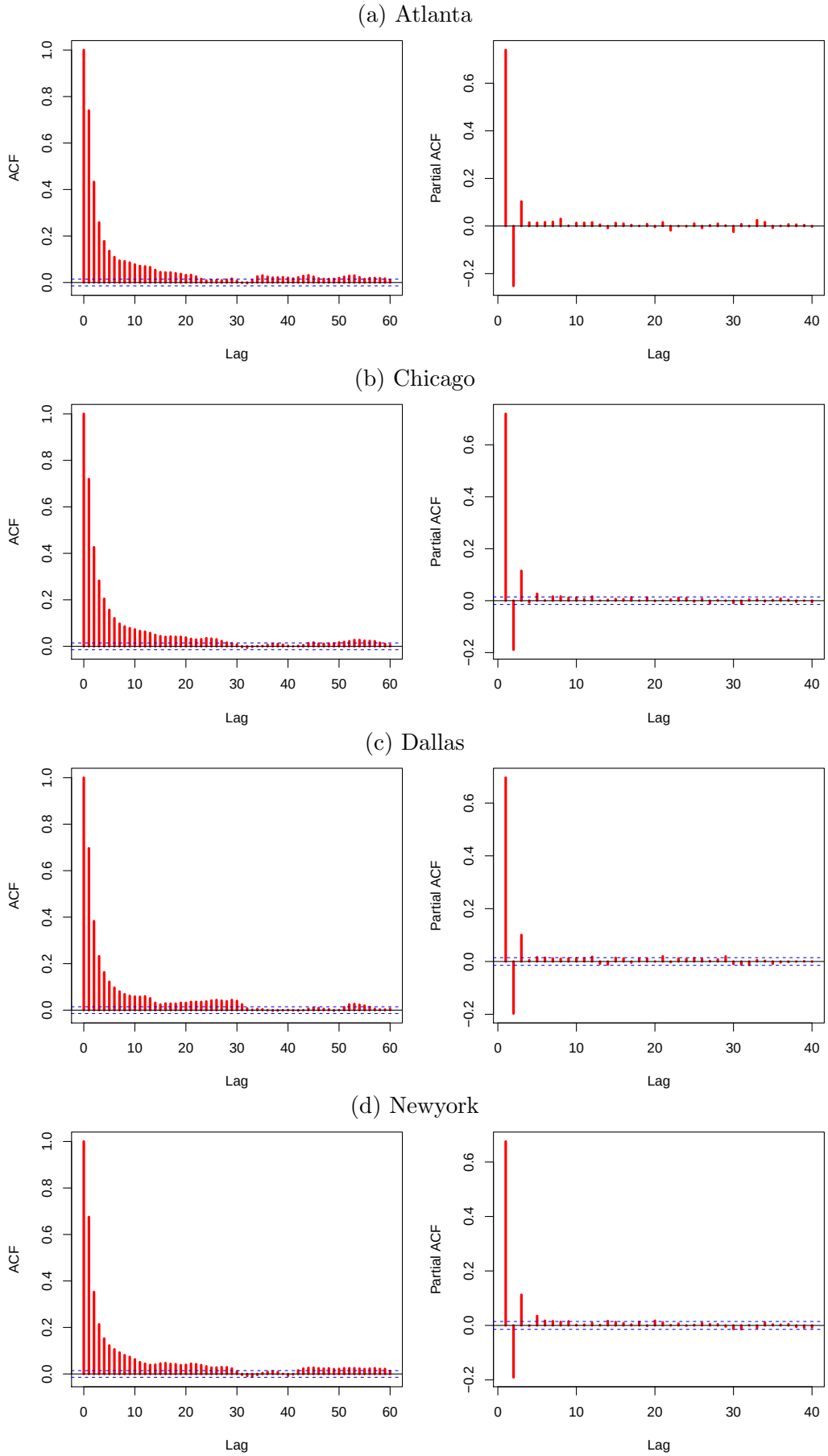



Figure 5.4: Sample ACFs and Sample PACFs of deseasonalized temperature data (part 1)

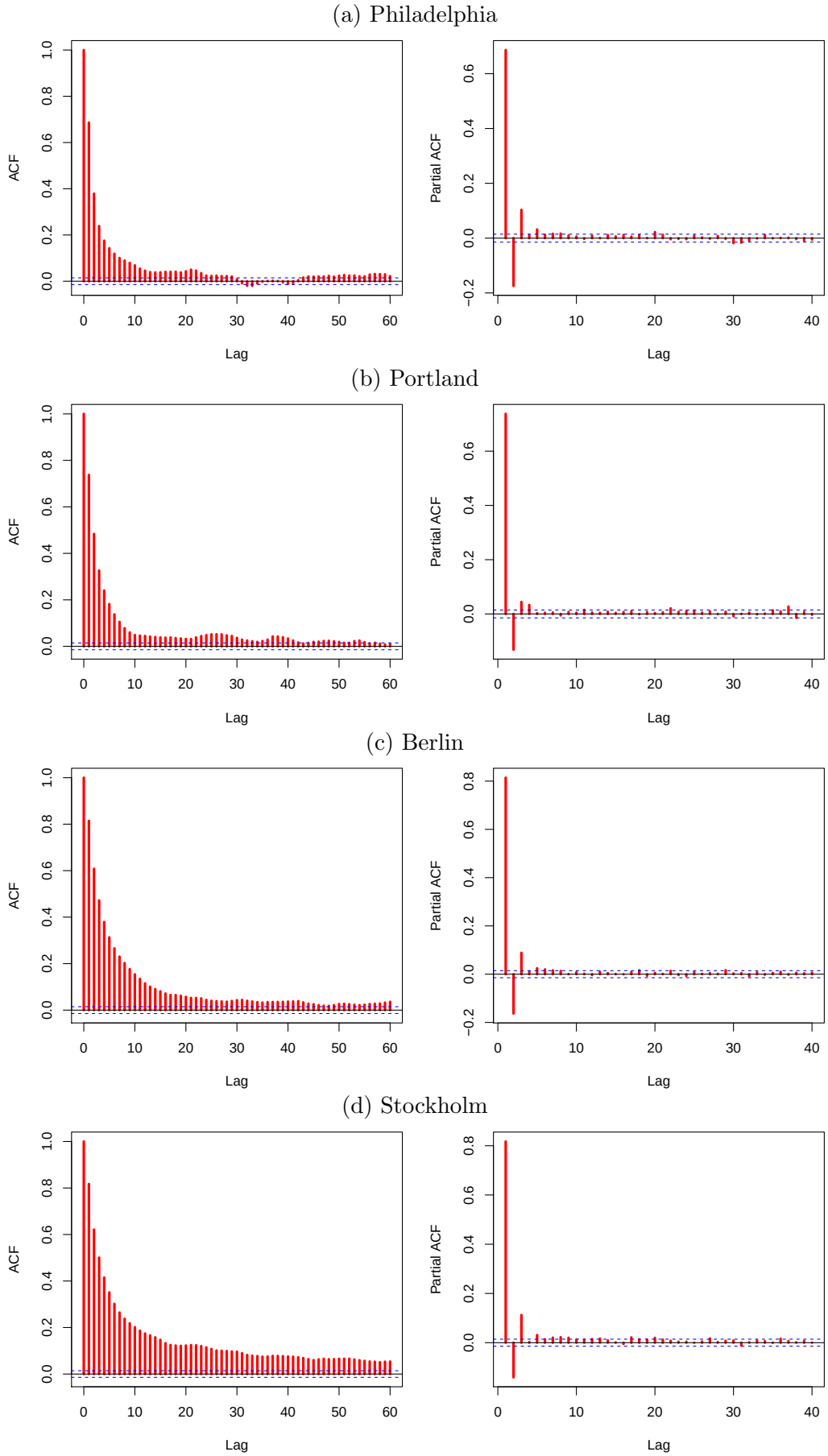



Figure 5.5: Sample ACFs and Sample PACFs of deseasonalized temperature data (part 2)

Table 5.6: Fitted parameters of AR(3)

| City | AR($p$) | | | | CAR($p$) | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | $\hat{b}_1$ | $\hat{b}_2$ | $\hat{b}_3$ | $\hat{b}_4$ | $\hat{a}_1$ | $\hat{a}_2$ | $\hat{a}_3$ | $\hat{a}_4$ | Real parts of eigenvalues of $\hat{\mathbf{A}}_{p\times p}$ |
| Atlanta | 0.9514 | -0.3471 | 0.1031 | - | 2.0486 | 1.4443 | 0.2926 | - | -0.34,-0.86,-0.86 |
| Chicago | 0.8766 | -0.2868 | 0.1148 | - | 2.1234 | 1.5336 | 0.2954 | - | -0.30,-0.91,-0.91 |
| Dallas | 0.8530 | -0.2807 | 0.1005 | - | 2.1470 | 1.5747 | 0.3272 | - | -0.34,-0.90,-0.90 |
| Newyork | 0.8255 | -0.2811 | 0.1124 | - | 2.1745 | 1.6301 | 0.3432 | - | -0.34,-0.92,-0.92 |
| Philadelphia | 0.8237 | -0.2573 | 0.1032 | - | 2.1763 | 1.6099 | 0.3304 | - | -0.33,-0.92,-0.92 |
| Berlin | 0.9281 | -0.2183 | 0.0851 | - | 2.0719 | 1.3621 | 0.2051 | - | -0.21,-0.93,-0.93 |
| Stockholm | 0.9481 | -0.2455 | 0.1124 | - | 2.0519 | 1.3493 | 0.1850 | - | -0.18,-0.93,-0.93 |
| Portland | 0.8388 | -0.1630 | 0.0164 | 0.0330 | 3.1612 | 3.6466 | 1.7930 | 0.2748 | -1.25,-0.26,-0.82,-0.82 |

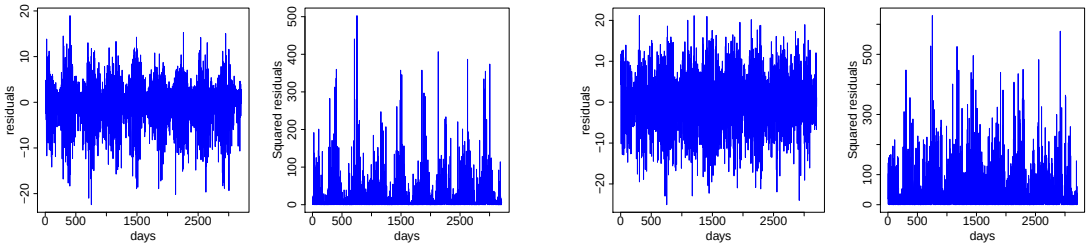
## 5.5 Seasonal Volatility

After fitting the deseasonalized temperature data to an AR($p$) process, we now have residuals $\hat{\epsilon}(t)$ from AR($p$). In order to give an overview of the residuals, Figure 5.6 plots the residuals $\hat{\epsilon}(t)$ and squared residuals $\hat{\epsilon}^2(t)$ of the first 9 years. We observe that the residuals are showing cyclic pattern with period of 1 year. To gain additional insight, Figure 5.7 plots the Sample ACFs of both the residuals and squared residuals. Seasonal pattern is observed in the Sample ACFs for squared residuals as well, which confirms the existence of seasonal variation.

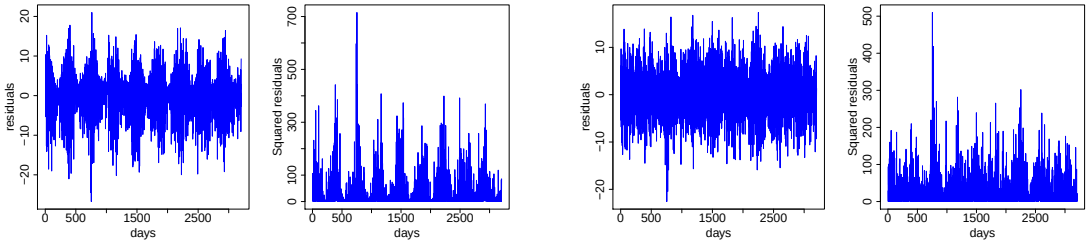
The residual $\epsilon(t)$ is has a multiplicative representation as in equations (3.23). In order to estimate $\sigma(t)$, we group them into 365 groups for each calendar day, and take the average of the squared residuals as the estimated variance for each calendar day. We use FTS as in equation (3.20) to capture seasonality in the variance. The parameters are estimated by least square minimization. Table 5.7 lists the fitted parameters for the FTS function. The number of periodic terms, $M$, is chosen by minimizing the BIC.

Figure 5.8 plots the fitted function over the empirical mean of squared residuals on each calendar day. We observe that the variance usually has larger value in winter and smaller value in summer. The pattern varies from city to city and depends on the location.

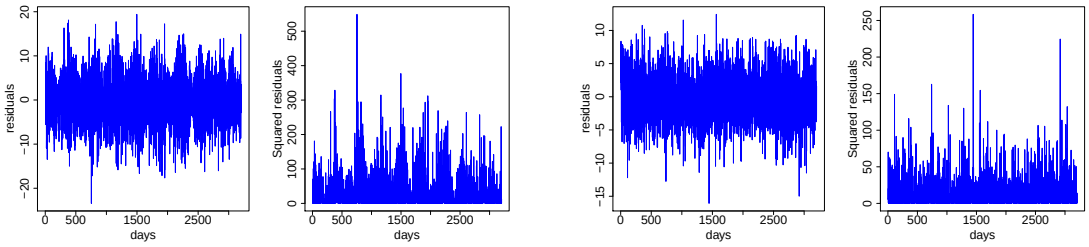
(a) Atlanta (b) Chicago




(c) Dallas (d) Newyork




(e) Philadelphia (f) Portland




(g) Berlin (h) Stockholm

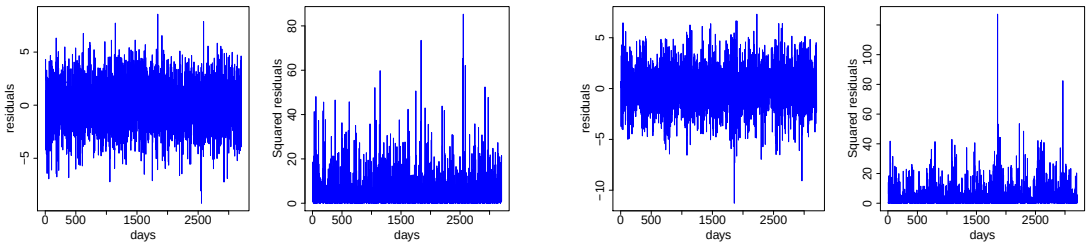



Figure 5.6: Residuals $\hat{\epsilon}(t)$ (left) and squared residuals $\hat{\epsilon}^2(t)$ (right)

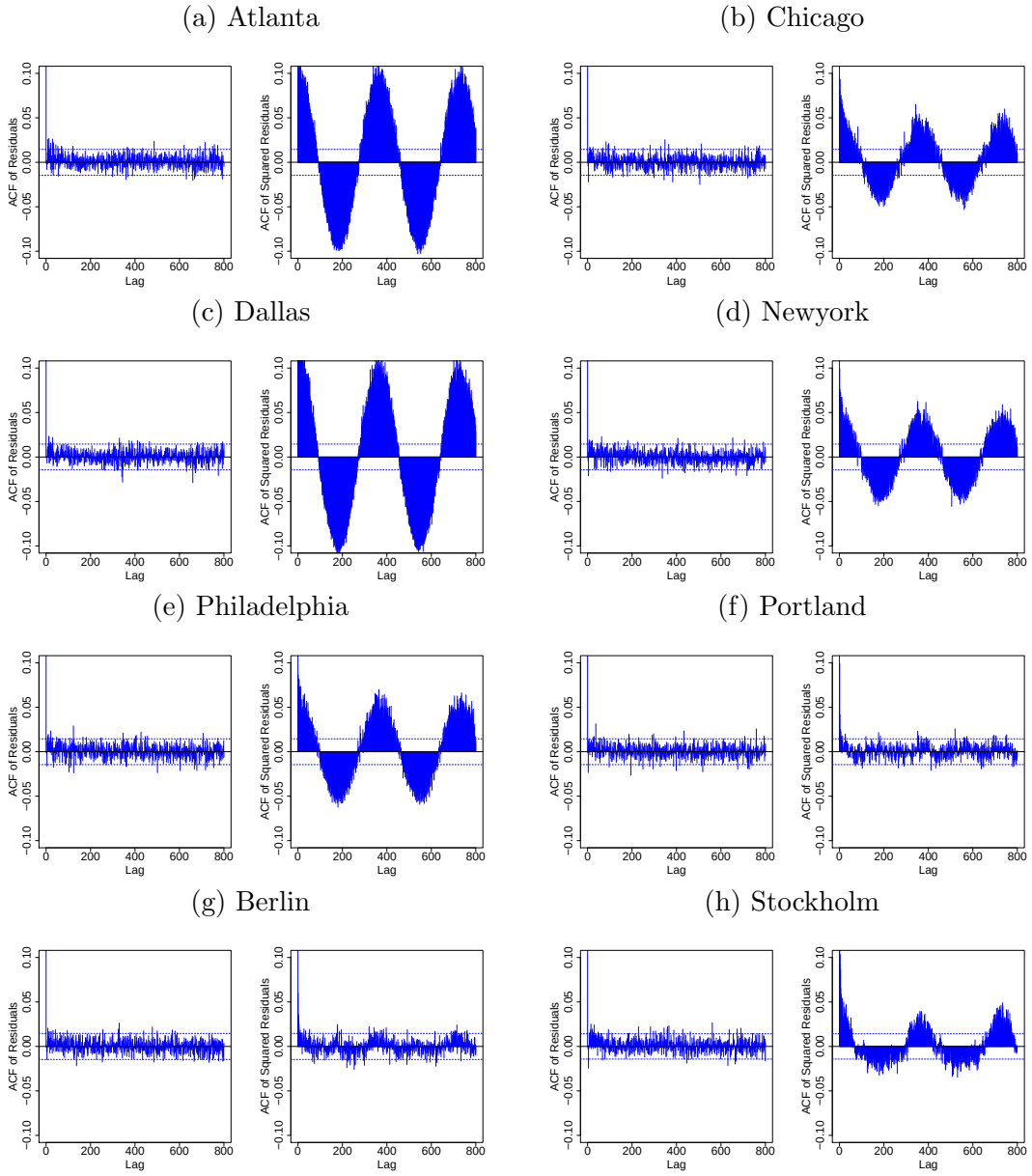



Figure 5.7: Sample ACFs of residuals $\hat{\epsilon}(t)$ (left) and Sample ACFs of squared residuals $\hat{\epsilon}^2(t)$ (right)

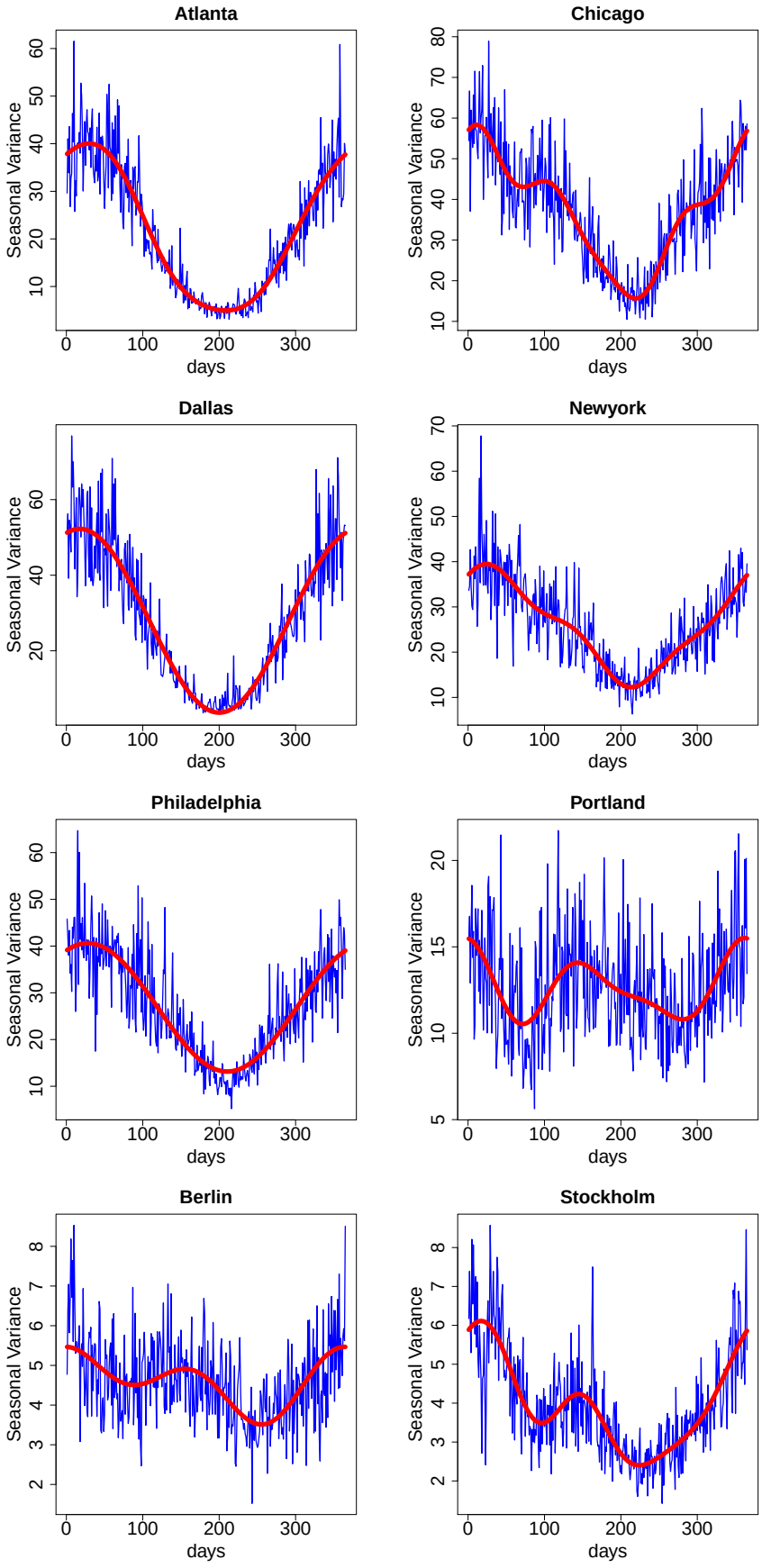



Figure 5.8: Average squared residuals (blue) and the fitted seasonal variance $\hat{\sigma}^2(t)$ (red)

.

Table 5.7: Fitted parameters of the seasonal variance function $\sigma^2(t)$

| City | $\hat{d}_1$ | $\hat{d}_2$ | $\hat{d}_3$ | $\hat{d}_4$ | $\hat{d}_5$ |
|---|---|---|---|---|---|
| Atlanta | 21.26(20.74,21.78) | 16.54(15.80,17.27) | 6.97( 6.24, 7.70) | - | - |
| Chicago | 37.24(36.48,38.01) | 15.13(14.06,16.21) | 7.42( 6.34, 8.49) | 0.13(-0.94, 1.20) | -2.90(-3.97,-1.82) |
| Dallas | 27.95(27.21,28.69) | 23.21(22.16,24.25) | 7.20( 6.15, 8.24) | - | - |
| Newyork | 25.73(25.16,26.29) | 9.85( 9.05,10.65) | 6.31( 5.51, 7.11) | 0.75(-0.03, 1.55) | -0.38(-1.17, 0.41) |
| Philadelphia | 26.88(26.25,27.51) | 12.16(11.27,13.05) | 6.38( 5.49, 7.27) | - | - |
| Portland | 12.62(12.33,12.92) | 0.38(-0.02, 0.80) | 0.30(-0.10, 0.72) | 1.57( 1.16, 1.99) | -0.74(-1.15,-0.32) |
| Berlin | 4.58( 4.48, 4.68) | 0.38( 0.25, 0.52) | 0.43( 0.29, 0.56) | 0.50( 0.36, 0.63) | -0.22(-0.35,-0.09) |
| Stockholm | 3.96( 3.87,4.04) | 1.15( 1.02,1.27) | 0.68( 0.55,0.80) | 0.67( 0.54,0.79) | -0.12(-0.24,0.00) |

| City | $\hat{d}_6$ | $\hat{d}_7$ | $\hat{d}_8$ | $\hat{d}_9$ | $\hat{M}$ | BIC |
|---|---|---|---|---|---|---|
| Atlanta | - | - | - | - | 1 | 2237 |
| Chicago | 2.30( 1.22, 3.37) | 2.34( 1.27, 3.42) | 2.04(1.00,3.09) | 1.56(0.51,2.60) | 4 | 2525 |
| Dallas | - | - | - | - | 1 | 2494 |
| Newyork | 0.68(-0.11, 1.47) | 1.82( 1.02, 2.62) | - | - | 3 | 2318 |
| Philadelphia | - | - | - | - | 1 | 2377 |
| Portland | 0.89( 0.49,1.30) | 0.11(-0.29,0.52) | - | - | 3 | 1822 |
| Berlin | - | - | - | - | 2 | 1023 |
| Stockholm | 0.14( 0.01,0.26) | 0.40( 0.27,0.52) | -0.05(-0.17,0.07) | -0.02(-0.14,0.10) | 4 | 975 |

The standardized residuals are calculated as $\hat{e}(t) = \hat{\epsilon}(t)/\hat{\sigma}(t)$. Figure 5.9 plots the Sample ACFs of $\hat{e}(t)$, $\hat{e}^2(t)$ and Sample PACFs of $\hat{e}(t)$. From Figure 5.9, the Sample ACFs of standardized residuals show that the correlations between the first few lags are significant. In Atlanta and Dallas the Sample ACFs of standardized residuals tail off. The Sample ACFs and Sample PACFs of squared standardized residuals shows significant dependency structure for the first few lags. An tails off behavior can to found in the Sample ACFs of standardized residuals for Chicago, Dallas and Stockholm and in the Sample PACF of standardized residuals for Berlin and Stockholm.

We perform the Box-Ljung test on the first few lags. Table 5.8 presents the Box-Ljung statistics for the first 1, 5 and 10 lags. For cities Atlanta, Dallas, Newyork and Stockholm, the test statistics for some lag are significant according to the p-value. Therefore, the hypothesis of randomness is rejected. For cities Berlin, Chicago and Portland, the test statistics are insignificant, and the null hypothesis of independence cannot be rejected under these number of lags.

Table 5.8: Results of Box-Ljung tests for the first few lags

| | lag =1 | | lag=5 | | lag=10 | |
|---|---|---|---|---|---|---|
| | $\hat{Q}_1$ | p-value | $\hat{Q}_5$ | p-value | $\hat{Q}_{10}$ | p-value |
| Atlanta | 7.5614 | 0.0059 | 42.8978 | 0.0000 | 90.4667 | 0.0000 |
| Berlin | 1.5576 | 0.2120 | 5.6679 | 0.3399 | 11.5207 | 0.3184 |
| Chicago | 2.7212 | 0.0990 | 6.6059 | 0.2516 | 13.3396 | 0.2053 |
| Dallas | 0.5311 | 0.4662 | 45.6039 | 0.0000 | 95.6145 | 0.0000 |
| Newyork | 1.2364 | 0.2662 | 9.405 | 0.0939 | 20.3647 | 0.0259 |
| Philadelphia | 0.0456 | 0.8309 | 6.3521 | 0.2734 | 18.1697 | 0.0521 |
| Portland | 0.9875 | 0.3204 | 1.2392 | 0.9411 | 3.752 | 0.9578 |
| Stockholm | 6.5142 | 0.0107 | 12.0081 | 0.0346 | 22.3863 | 0.0132 |

$\hat{Q}_m$ is the Ljung-Box Q statistics with lags $m$.

A dependence structure exists to be explored in the standardized residuals. However, dependencies in the standardized residuals are not accounted for in our model. An usual approach to cope with this is to extend the model by adding an generalized autoregressive conditional heteroskedasticity (GARCH) term to the volatility function $\sigma^2(t)$, see for example, Härdle and Cabrera [2012]. One can use a more sophisticated model with stochastic volatility, which is outside the scope of this thesis.

## 5.6 Distribution of the Random Noise

Table 5.9 lists some descriptive statistics of the standardized residuals. Most cities have negative skewness and positive excessive kurtosis. The hypothesis of normality is rejected for all cities under the Jarque-Bera (JB) test.

Table 5.9: Descriptive statistics of the standardized residuals

| | Mean | Std. dev | Max | Median | Min | Skewness | Kurtosis | JB stat |
|---|---|---|---|---|---|---|---|---|
| Atlanta | 0.00 | 1.02 | 4.11 | 0.13 | -6.44 | -0.64 | 0.90 | 1862 |
| Chicago | -0.00 | 1.00 | 4.04 | 0.03 | -4.10 | -0.15 | 0.28 | 127 |
| Dallas | -0.00 | 1.02 | 4.03 | 0.11 | -8.19 | -0.68 | 1.45 | 2999 |
| Newyork | -0.00 | 1.00 | 3.66 | 0.02 | -4.17 | -0.11 | 0.09 | 45 |
| Philadelphia | 0.00 | 1.00 | 3.43 | 0.03 | -4.18 | -0.19 | 0.13 | 117 |
| Portland | 0.00 | 1.00 | 3.77 | -0.02 | -4.57 | 0.01 | 0.22 | 36 |
| Berlin | 0.00 | 1.00 | 3.90 | -0.01 | -4.60 | 0.04 | 0.24 | 49 |
| Stockholm | -0.00 | 1.00 | 3.88 | 0.01 | -4.68 | -0.09 | 0.34 | 115 |

*St. dev* denotes the standard deviation , while *JB stat* denotes the Jarque-Bera statistic. The critical values for JB test are 5.99 at 5% significance level, and 9.21 at 1% significance level.

Motivated by the existence of negative skewness and heavy tails, we fit the standardized residuals with GHYP distribution and two of its limiting cases: the HYP distribution and the NIG distribution. Package "ghyp" in R is used to estimate parameters. Estimated parameters are shown in Table 5.10.

In order to examine and compare the goodness-of-fit for the four distributions: Gaussian, GHYP, HYP and NIG, we perform the Kolmogorov-Smirnov (K-S) test and Anderson-Darling (A-D) test, as well as the QQ plot.

The results for the K-S test and the A-D test are shown in Table 5.11. Recall that the null hypothesis of the A-D test and K-S test is that the data comes from the hypothesized distribution. We reject the null hypothesis when the test statistic is larger than the critical values in both tests. Therefore, we have that the null hypothesis of normal distribution is rejected at confidence level of 5% for almost all cities in both tests, with exception for Berlin in the K-S test. Since the critical values are not available for GHYP, HYP and NIG distributions, we rely on the K-S test for them. The null hypotheses that the data has GHYP distribution and that the data has HYP distribution can not be rejected at the 5% confidence level in the K-S test for all cities. The null hypothesis that the data comes from the NIG distribution can not be rejected at confidence level of 5% for most cities, with the exception of Atlanta and Stockholm. The K-S test results indicate that we have no sufficient evidence to reject the hypotheses /HYof GHYP, NIG and HYP distributions

| | | $\hat{\lambda}$ | $\hat{\alpha}$ | $\hat{\beta}$ | $\hat{\delta}$ | $\hat{\mu}$ |
|---|---|---|---|---|---|---|
| Atlanta | GHYP | 3.026 | 2.750 | -0.889 | 0.082 | 0.796 |
| | HYP | - | 2.473 | -1.000 | 1.289 | 0.883 |
| | NIG | - | 2.368 | -1.127 | 1.812 | 0.981 |
| Chicago | GHYP | 4.819 | 3.516 | -0.419 | 1.440 | 0.414 |
| | HYP | - | 3.040 | -0.438 | 2.459 | 0.433 |
| | NIG | - | 2.846 | -0.449 | 2.777 | 0.443 |
| Dallas | GHYP | 2.996 | 2.656 | -0.740 | 0.047 | 0.682 |
| | HYP | - | 2.212 | -0.756 | 1.167 | 0.705 |
| | NIG | - | 1.977 | -0.810 | 1.673 | 0.751 |
| Newyork | GHYP | 9.150 | 6.605 | -1.260 | 4.636 | 1.229 |
| | HYP | - | 6.084 | -1.318 | 5.534 | 1.284 |
| | NIG | - | 5.964 | -1.303 | 5.676 | 1.271 |
| Portland | GHYP | 4.661 | 4.300 | 0.152 | 2.912 | -0.151 |
| | HYP | - | 3.868 | 0.141 | 3.467 | -0.140 |
| | NIG | - | 17.995 | -1.947 | 17.784 | 1.936 |
| Philadelphia | GHYP | 6.375 | 5.634 | -1.482 | 3.745 | 1.412 |
| | HYP | - | 5.313 | -1.540 | 4.534 | 1.463 |
| | NIG | - | 5.210 | -1.560 | 4.722 | 1.482 |
| Berlin | GHYP | 3.913 | 4.127 | 0.365 | 2.876 | -0.363 |
| | HYP | - | 3.684 | 0.221 | 3.255 | -0.220 |
| | NIG | - | 3.482 | 0.220 | 3.469 | -0.219 |
| Stockholm | GHYP | 3.529 | 3.429 | -0.215 | 2.018 | 0.215 |
| | HYP | - | 3.075 | -0.220 | 2.555 | 0.220 |
| | NIG | - | 2.849 | -0.221 | 2.832 | 0.221 |

Table 5.10: Estimated parameters for the generalized hyperbolic distributions, hyperbilic distributions and normal inverse Gaussian distribution

for most cities.

Comparing the test statistics of the four distributions in both tests, the values in the test of Normal distribution are the largest and lead to rejection of the null hypothesis. The test statistics in the test of the GHYP distribution usually provides the smallest values which indicates a better fit. This is expected since the GHYP distribution is the superclass of other distributions. The test statistics for HYP are smaller than those for NIG, except for Berlin. Hence we have the general conclusion that the GHYP distribution provides the best fit in four distributions, the HYP distribution the second, NIG distribution the third and the normal distribution the worst.

Figure 5.10 presents the QQ plots of the data versus the four theoretical distributions. For normal distribution, the quantiles deviate from the 45-degree reference line in the tails apparently in cities Atlanta, Dallas and Stockholm. This shows that the normal

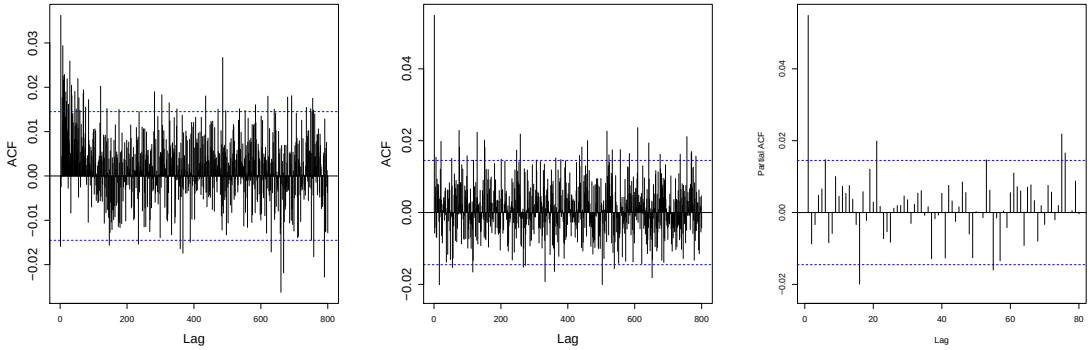
Table 5.11: Testing distributions of the standardized residuals

| City | | Normal | GHYP | NIG | HYP | 5% CV | 1% CV |
|---|---|---|---|---|---|---|---|
| Atlanta | K-S | 0.058 | 0.0093 | 0.0115 | 0.0108 | 0.0101 | 0.0121 |
| | A-D | 115.1 | 2.921 | 4.8346 | 3.925 | - | - |
| Berlin | K-S | 0.0083 | 0.0053 | 0.0053 | 0.0053 | 0.0102 | 0.0122 |
| | A-D | 2.5718 | 0.5322 | 0.5286 | 0.5257 | - | - |
| Chicago | K-S | 0.023 | 0.008 | 0.0092 | 0.0087 | 0.0101 | 0.0121 |
| | A-D | 13.71 | 1.349 | 1.7117 | 1.5967 | - | - |
| Dallas | K-S | 0.0533 | 0.0047 | 0.0067 | 0.0052 | 0.0101 | 0.0121 |
| | A-D | 97.4627 | 0.3618 | 1.2572 | 0.6903 | - | - |
| Newyork | K-S | 0.0121 | 0.0041 | 0.0042 | 0.0041 | 0.0101 | 0.0121 |
| | A-D | 4.0394 | 0.4327 | 0.4558 | 0.4477 | - | - |
| Philadelphia | K-S | 0.0172 | 0.0046 | 0.0048 | 0.0047 | 0.0101 | 0.0121 |
| | A-D | 9.4205 | 0.5459 | 0.5786 | 0.5706 | - | - |
| Portland | K-S | 0.0123 | 0.0076 | 0.0133 | 0.0078 | 0.0101 | 0.0121 |
| | A-D | 9.8496 | 1.761 | 4.4689 | 1.7877 | - | - |
| Stockholm | K-S | 0.0122 | 0.0046 | 0.0049 | 0.0048 | 0.0099 | 0.0118 |
| | A-D | 6.1018 | 0.3263 | 0.3656 | 0.3517 | - | - |

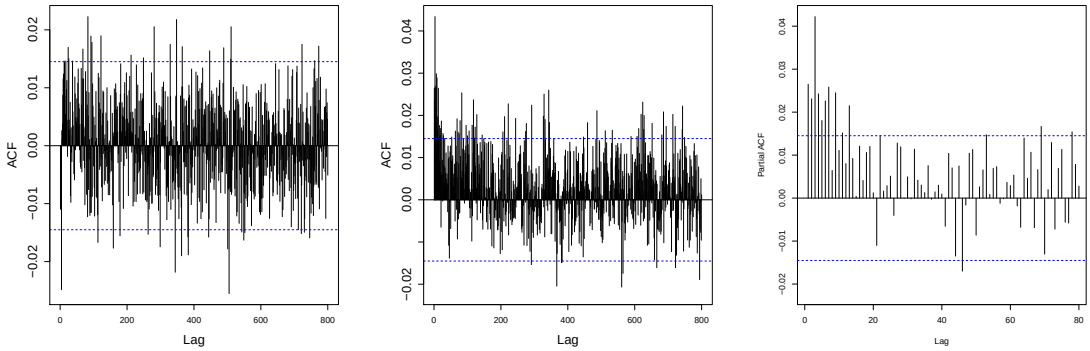
*CV* denotes the critical value. *K-S* denotes Kolmogorov distance, while *A-D* denotes the Anderson-Darling statistic. Critical values for A-D statistic are available for normal distribution, but not for the HYP, NIG and GHYP distributions. For A-D test for normal distribution, the 5% CV is 2.492, the 1% CV is 3.857.

distribution provides a poor fit. The quantiles for GHYP, HYP and NIG distributions are much closer to reference line in cities Dallas, Philadelphia, Portland, Berlin. In most cities, the quantiles for GHYP, HYP and NIG are close to each other and they often overlap. We only observe in the plot of Atlanta that quantiles for GHYP are closer to the reference line compared to others.

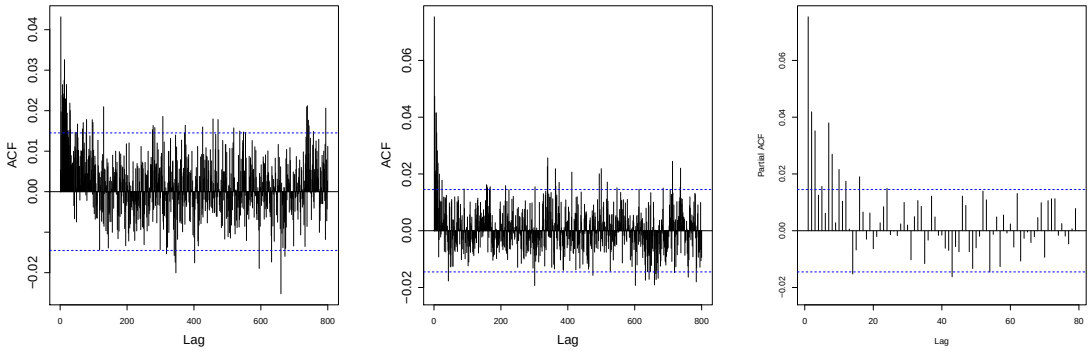
(a) Atlanta




(b) Chicago




(c) Dallas




(d) Newyork

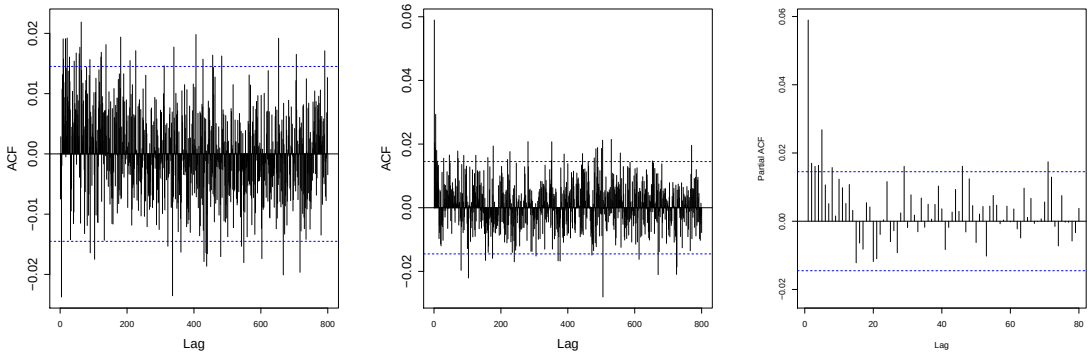
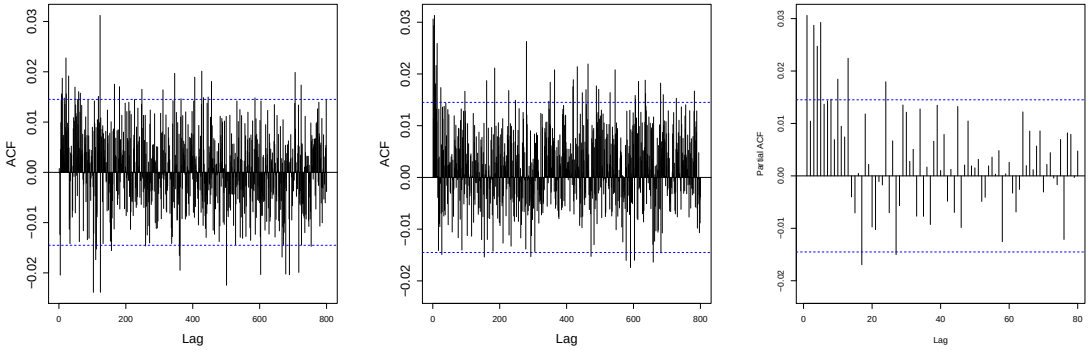
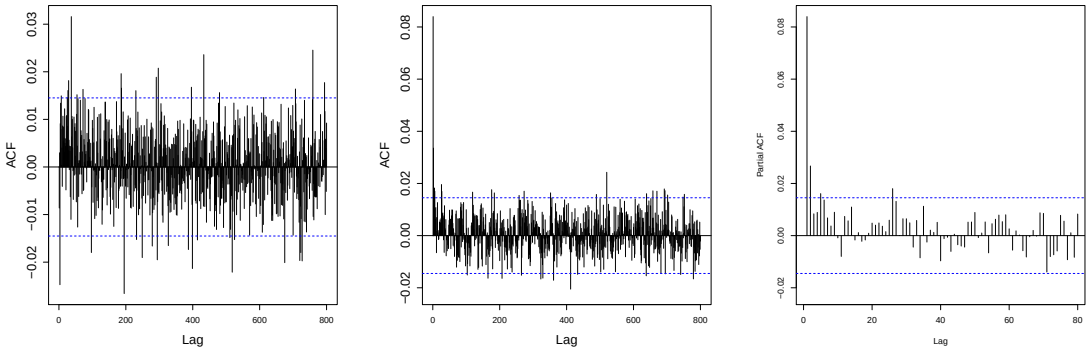



Figure 5.9: Sample ACFs of standardized residuals (Left), Sample ACFs of squared standardized residuals (middle), and Sample PACFs of standardized residuals (Right)

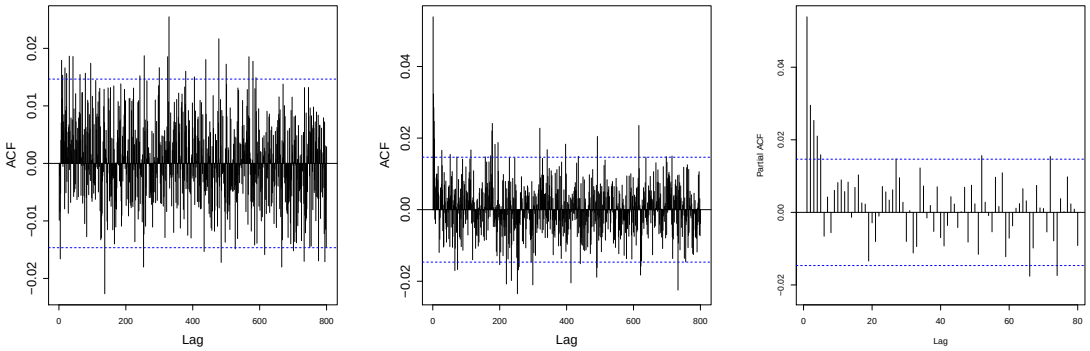
(e) Philadelphia




(f) Portland




(g) Berlin




(h) Stockholm

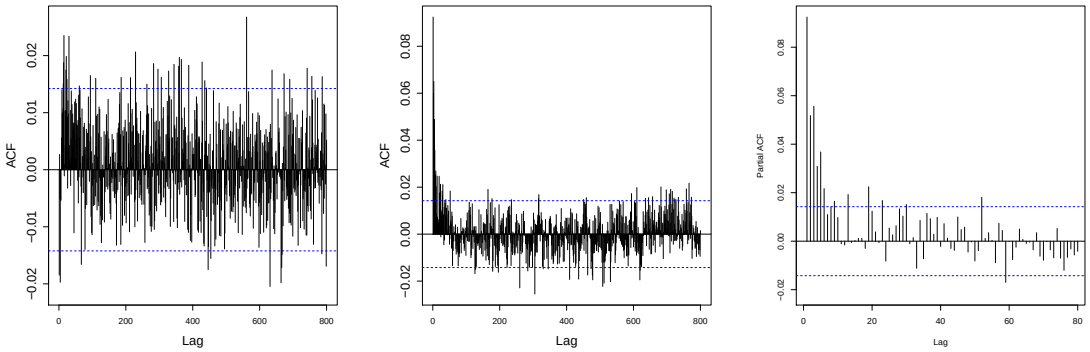
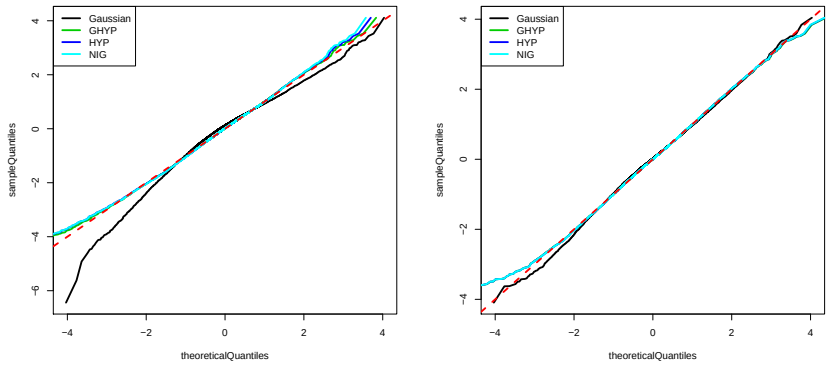
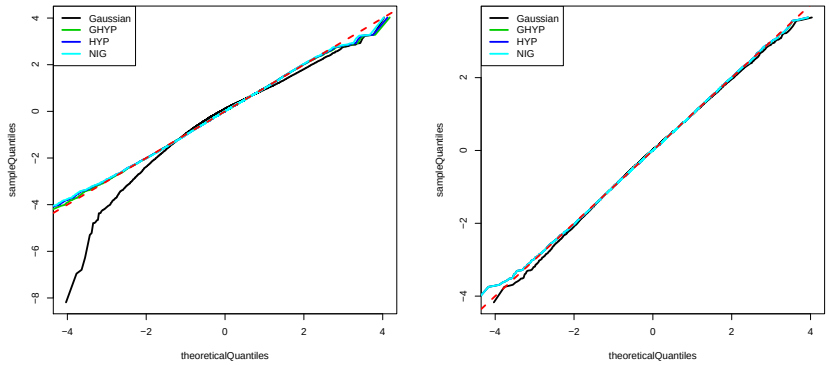



Figure 5.9: Sample ACFs of standardized residuals (Left), Sample ACFs of squared standardized residuals (middle), and Sample PACFs of standardized residuals (Right)

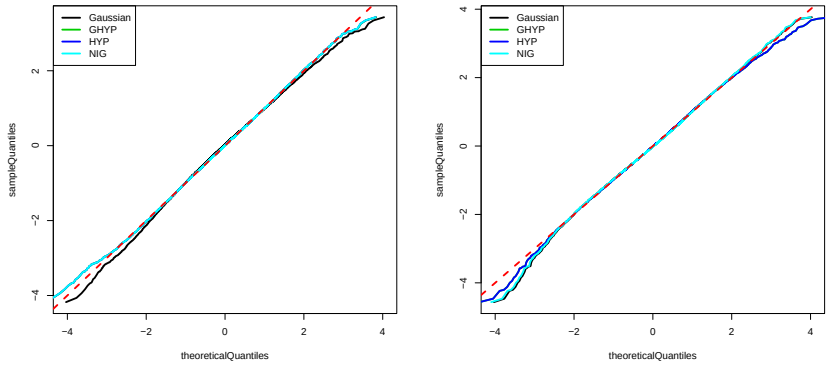
(a) Atlanta (b) Chicago




(c) Dallas (d) Newyork




(e) Philadelphia (f) Portland




(g) Berlin (h) Stockholm

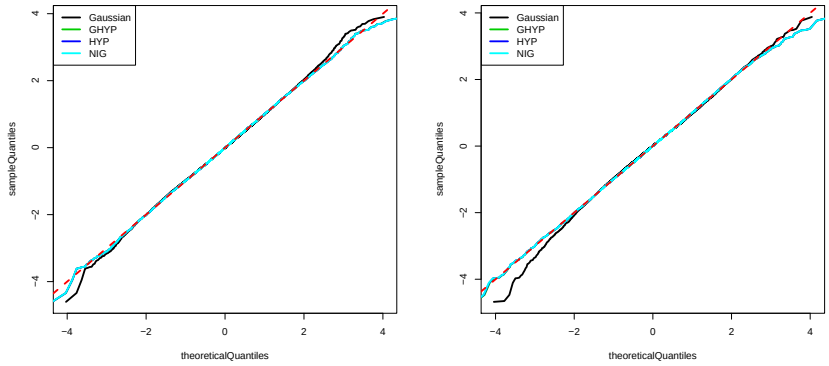



Figure 5.10: QQ plot of standardized residuals

# Chapter 6

# Empirical Results of Pricing

With the fitted results of the temperature data from previous chapter of data analysis, we now implement the pricing framework for CAT futures and HDD futures for Stockholm. Empirical results on CAT futures are presented in first section. Results on HDD futures are shown in the second section.

## 6.1 Empirical Results for CAT Futures Prices

We consider 1195 contracts with 8 different measurement periods traded during 11/20/2011 $\sim$ 9/11/2012 on Stockholm CAT futures.

Recall from Theorem that 4.6, the CAT futures price is given by

$$F_{CAT}(t,\tau_1,\tau_2) = \int_{\tau_1}^{\tau_2} \Lambda(u)du + \mathbf{a}(t,\tau_1,\tau_2)\mathbf{X}(t) + \int_t^{\tau_1} \sigma(u)\psi'(\theta(u))\mathbf{a}(t,\tau_1,\tau_2)\mathbf{e}_p du + \int_{\tau_1}^{\tau_2} \sigma(u)\psi'(\theta(u))\mathbf{e}_1^\top \mathbf{A}^{-1}\left(e^{\mathbf{A}(\tau_2-u)} - I_p\right)\mathbf{e}_p du,$$

where $\mathbf{a}(t,\tau_1,\tau_2) = \mathbf{e}_1^\top \mathbf{A}^{-1}\left(e^{\mathbf{A}(\tau_2-t)} - e^{\mathbf{A}(\tau_1-t)}\right)$; $I_p$ is a $p\times p$ identity matrix; $\psi(u)$ is the logarithm of the moment generating function of $L(1)$; $\theta = \theta(u)$ is the market price of risk (MPR).

By previous Chapter of data analysis, we have estimated $\hat{\Lambda}(t)$ and $\hat{\mathbf{A}}$, $\hat{\sigma}(t)$. With the fitted parameters for different marginal distributions of $L$, we have $\psi'_{BM}(u)$, $\psi'_{GHYP}(u)$, $\psi'_{HYP}(u)$ and $\psi'_{NIG}(u)$ respectively, see Figure 6.1 for the plot of these functions. $p = 3$ is chosen for most cities, then $\hat{\mathbf{X}}(t)$ is a 3-dimensional vector given by

$$\hat{\mathbf{X}}(t) = \left(\hat{X}_1(t), \hat{X}_1(t+1) - \hat{X}_1(t), \hat{X}_1(t+2) - 2\hat{X}_1(t+1) + \hat{X}_1(t)\right)^\top, \qquad (6.1)$$

where $\hat{X}_1(t) = T(t) - \hat{\Lambda}(t)$. We see that although the price is calculated with information up to $t$, we requires information of at time $t+1$ and $t+2$. Since we perform our analysis in sample, we have temperature data at time $t+1$ and $t+2$ and we use them directly. For out of sample evaluation, one can resort to temperature forecast. For more discussion on recovering the state vector $\mathbf{X}(t)$, one can refer to Benth and Benth [2012].

In order to study the significance of the market price of risk $\theta$, we first examine the CAT futures prices estimated from the formula assuming that the market price of risk $\theta(u) \equiv 0$, such prices are denoted by $F_{CAT,0}$. Figure 6.2 plots the observed CAT futures prices, the realized CAT values and the theoretical prices $F_{CAT,0}$ for Stockholm. The observed CAT futures prices in the market remain almost unchanged, which indicates the market is not liquid in reality. The theoretical prices $F_{CAT,0}$. is smaller than the observed futures prices for all contracts. The realized CAT lies in the middle of the market price and the theoretical prices. The theoretical prices when $\theta(u) \equiv 0$ in the cases of normal, GHYP, HYP and NIG distributions are the same. This agrees with the fact that when $\theta = 0$, $Q = P$ and the price is calculated under unique the physical measure $P$. $\psi'(0)$ is almost zero, see Figure 6.1. Therefore, the CAT futures price is determined by the first two terms in the formula, which is the integrated seasonal mean and the effect from temperature variations.

From Figure 6.2, the theoretical prices are constant at close dates far from the measurement period, and become varying when close to the measurement period. Since the eigenvalues of $\mathbf{A}$ all have negative real parts, $e^{\mathbf{A}(\tau_2 - t)}$ and $e^{\mathbf{A}(\tau_1 - t)}$ both converge to zero as $\tau_1 - t \to \infty$, hence the second term $\mathbf{a}(t, \tau_1, \tau_2)\mathbf{X}(t)$ in the formula converges to zero given $\mathbf{X}(t)$. The CAT futures price is then equal to the integrated seasonal mean over the measurement period when far from measurement period.

We can further investigate the contributions of the first two terms in the formula by empirical data. Take the CAT futures contract with monthly measurement period of July 2012 as an example. The first term $\int_{\tau_1}^{\tau_2} \Lambda(u)du$, which is the integrated seasonal mean on the measurement period, gives the constant value of 522. On the close dates far from the begin of the measurement period, $\mathbf{a}(t, \tau_1, \tau_2)$ is nearly zero. The CAT futures price is determined by the integrated seasonal mean which is constant. However, as the close date becomes closer to $\tau_1$, $\mathbf{a}(t, \tau_1, \tau_2)$ becomes large enough to affect the price. When the close date is 6 days prior to June 1, that is $\tau_1 - t = 6$, $\hat{\mathbf{a}}(t, \tau_1, \tau_2) = (2.6, 4.8, 2.6)$, $\hat{\mathbf{X}}(t) = (0.1, -1.5, 0.2)$ the second term has the value of -14, which is about 3% of the first term. When the close date is 2 days prior to June 1, $\hat{\mathbf{a}}(t, \tau_1, \tau_2) = (6.3, 10.6, 5.3)$, and $\hat{\mathbf{X}}(t) = (-1.4, -1.4, 0.2)$, the second term has the value of -20, which is about 4% of the first term. As the close date gets closer to the start of measurement period, the second

term in CAT futures price, referred as *the effect of the temperature variations* in Benth and Šaltytė Benth [2011], impacts the CAT futures price more significantly. However, even when close to the measurement period, the value of the integrated seasonal mean still has a larger contributions to the futures price compared to the effect of the seasonal variation. And overall the seasonal mean determines the level of the CAT futures price.

For contracts with longer measurement period, for example the seasonal strip in Figure 6.2, the integrated seasonal mean is larger compared to the second term. The price is then less sensitive to the temperature variations when close to measurement period.

The deviation of theoretical prices $F_{CAT,0}$ from observed prices indicates the existence of non-zero MPRs. We restrict to the assumption that the MPR is constant over $[t, \tau_2^i]$ for each contract $i = 1, \ldots, I$, i.e. $\theta(u) = \theta_t^i(u)$. Recall that the MPR is calibrated by square minimization,

$$\begin{aligned}\hat{\theta}_{t,CAT}^i = \arg\min_{\theta_t^i} \Bigg( & F_{CAT}(t,\tau_1,\tau_2) - \int_{\tau_1}^{\tau_2} \hat{\Lambda}(u)du - \mathbf{a}(t,\tau_1,\tau_2)\mathbf{X}(t) \\ & -\psi'(\theta_t^i)\left\{ \int_t^{\tau_1} \hat{\sigma}(u)\mathbf{a}(t,\tau_1,\tau_2)\mathbf{e}_p du \right. \\ & \left. + \int_{\tau_1}^{\tau_2} \hat{\sigma}(u)\mathbf{e}_1^\top \mathbf{A}^{-1}\left(e^{\mathbf{A}(\tau_2-u)} - I_p\right)\mathbf{e}_p du \right\} \Bigg)^2 .\end{aligned}$$

Figure 6.3 plots the implied MPR for contracts with 8 different measurement periods. We see that the MPR in around 0.2. From Figure 6.1, we see that $\psi'(u)$ is nearly the same for the four different distributions we consider, i.e. the normal, GHYP, HYP and NIG distributions. Therefore, the implied constant MPR $\theta$ will be almost the same for four distributions which leads to the same $\psi'(\theta)$. Figure 6.4 plots the implied MPR with respect to Normal, NIG, HYP and GHYP distributions for contract with measurement period of September 2012, where the MPR are completely overlapping.

The last two terms in the CAT futures price, with $\psi'(0)$ being zero, is then the risk premium we define before, i.e.

$$RP = \int_t^{\tau_1} \sigma(u)\psi'(\theta(u))\mathbf{a}(t,\tau_1,\tau_2)\mathbf{e}_p du + \int_{\tau_1}^{\tau_2} \sigma(u)\psi'(\theta(u))\mathbf{e}_1^\top \mathbf{A}^{-1}\left(e^{\mathbf{A}(\tau_2-u)} - I_p\right)\mathbf{e}_p du,$$

Hence we notice that the risk premium is determined by the seasonal variation, and autocorrelation structure of the temperature and the market price of risk. Given different level of MPR, the RP term will contribute to the CAT futures differently. From the implied MPR result we obtain before, the MPR is around 0.2, which agrees with the results in Härdle and Cabrera [2012] and Benth et al. [2011]. Therefore, the RP term contributes

less significantly compared to the seasonal function.

If we use the implied MPR which is assumed to be constant to calculate the CAT futures price for our contracts, the CAT futures price simply is the same with the true market value.

## 6.2 Empirical Results HDD Futures Prices

From Theorem 4.10 the price of the HDD futures in the special case when $L$ is Brownian motion

$$F_{HDD}(t,\tau_1,\tau_2) = \int_{\tau_1}^{\tau_2} \nu(t,s)\Psi\left[\frac{c-m(t,s))}{\nu(t,s)}\right] ds,$$

where

$$m(t,u) = \Lambda(u) + \mathbf{e}_1^\top e^{\mathbf{A}(u-t)}\mathbf{X}(t) + \int_t^u \theta(s)\sigma(s)\mathbf{e}_1^\top e^{\mathbf{A}(u-s)}\mathbf{e}_p ds,$$

$$\nu^2(t,u) = \int_t^u \sigma^2(s)\left(\mathbf{e}_1^\top e^{\mathbf{A}(u-s)}\mathbf{e}_p\right)^2 ds.$$

$\Phi$ is the cumulative standard normal distribution function and $\Psi(x) = x\Phi(x) + \Phi'(x)$.

For general Lévy processes, the price of a HDD futures is estimated by Fourier transform, and numerically given by

$$F_{HDD}(t,\tau_1,\tau_2) \approx \sum_{u=\tau 1}^{\tau 2} \sum_{x_k \le d(u)} \left[\left(c - \Lambda(u) - \mathbf{e}_1^\top e^{\mathbf{A}(u-t)}\mathbf{X}(t)\right) - x_k\right] f_{G(u)}(x_k)\Delta x \Delta u$$

where $d(u) = c - \Lambda(u) - \mathbf{e}_1^\top e^{\mathbf{A}(u-t)}\mathbf{X}(t)$, $f(x_k,u)$ is determined by characteristic function

$$\phi_{G(u)}(\lambda) = \exp\left\{\int_t^u \psi\left(i\lambda\mathbf{e}_1^\top e^{\mathbf{A}(u-s)}\mathbf{e}_p\sigma(s) + \theta(s)\right) ds - \int_t^u \psi\left(\theta(s)\right) ds\right\},$$

with $\psi(u)$ being the log-moment generating function of $L(1)$.

Here we examine the HDD futures prices if the market price of risk $\theta(u) \equiv 0$. Furthermore, we only consider the NIG($\alpha$,$\beta$,$\mu$,$\delta$) distribution since $\psi(u)$ has simple expressions,

$$\psi_{NIG}(u) = \left\{\mu u + \delta\left(\sqrt{\alpha^2-\beta^2} - \sqrt{\alpha^2-(\beta+u)^2}\right)\right\}. \tag{6.2}$$

For the normal distribution, the price is calculated by closed form formula. For NIG distribution, the Fourier method is applied. Figure 6.5 plots the real market HDD futures prices , the realized HDD values and the theoretical prices given for the normal distribution and NIG distribution with $\theta(u) \equiv 0$ for Stockholm. Again, we notice that the HDD futures

prices remain almost constant in real market. We see that the theoretical prices when $\theta(u) \equiv 0$ in the cases of normal and NIG distributions are the same. Therefore we believe that the Fourier technique returns a plausible result. We also notice that the HDD futures price is constant when far from the start of the measurement period, and wiggles when approaching the start of the measurement period.

We analyze the contributions of each terms in the explicit formula under the case of Brownian motion. A HDD future with measurement January 2012 of Stockholm is studied. Let $u$ be the middle of the measurement period $[\tau_1, \tau_2]$. The seasonal mean $\hat{\Lambda}(u)$ is -0.918. When the close date $t$ is a month before $\tau_1$, we have $\mathbf{e}_1^\top e^{\mathbf{A}(u-t)}\mathbf{X}(t)$ equals to 0.002, which is about 2% of the $\hat{\Lambda}(u)$. When $t$ is 6 days before $\tau_1$, $\mathbf{e}_1^\top e^{\mathbf{A}(u-t)}\mathbf{X}(t)$ returns a value of 0.259, which is around 28% of $\hat{\Lambda}(u)$. Therefore, we have same conclusion that the effect of the temperature variations impacts the CAT futures price significantly when approaching the start of measurement, but the mean level of the price is set by the seasonal mean of the temperature.

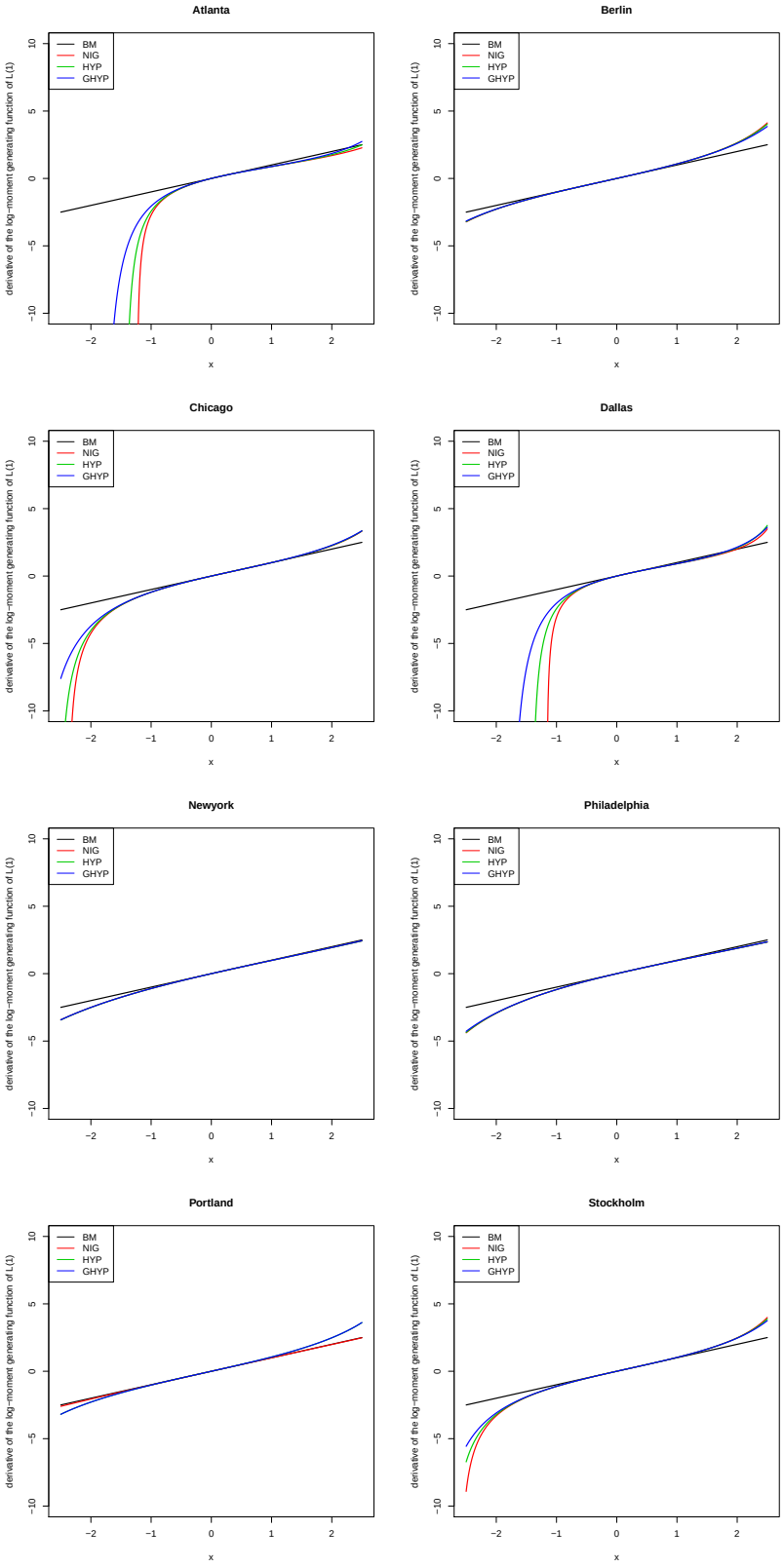



Figure 6.1: The derivative of the log-moment generating function

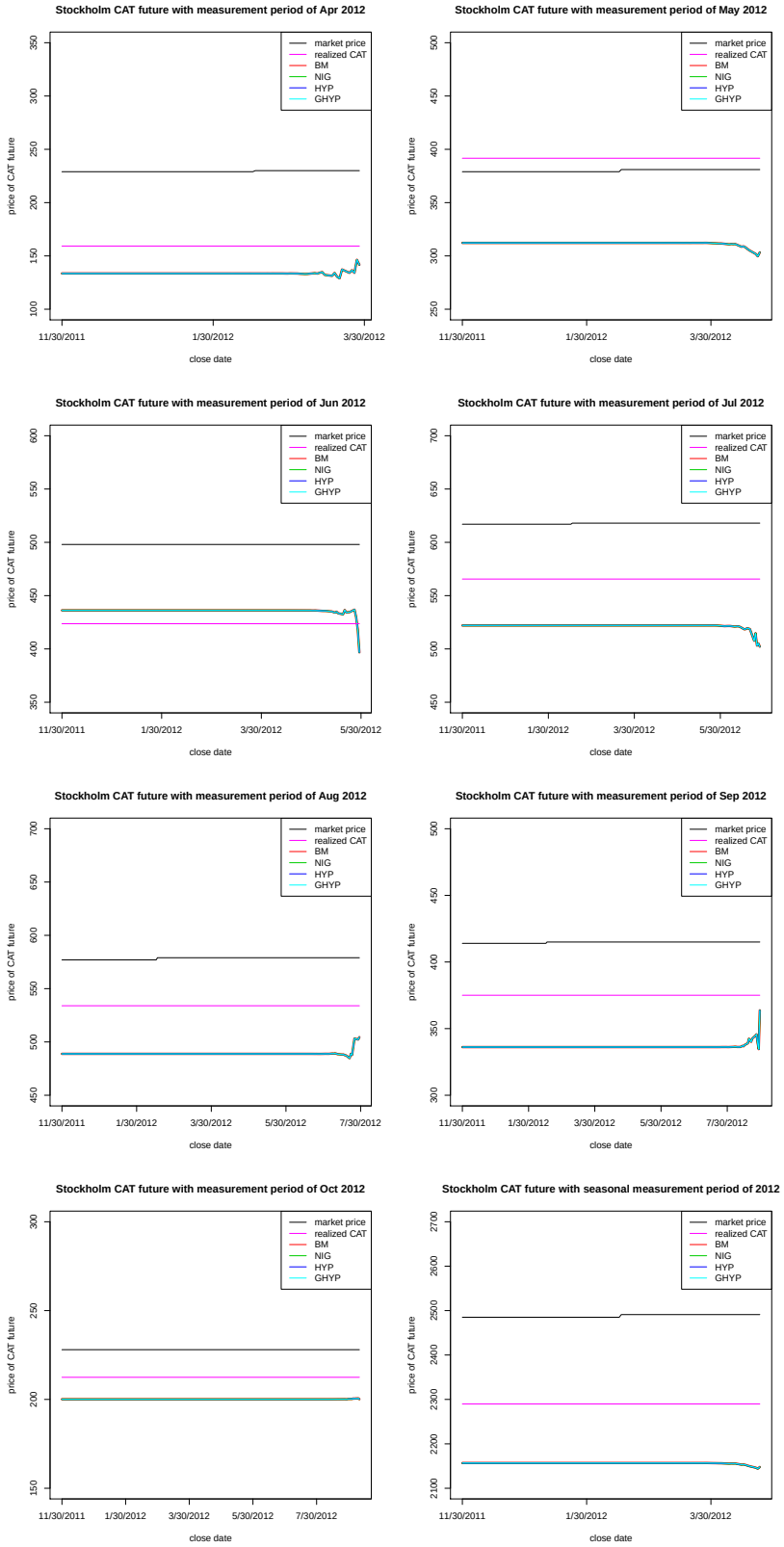



Figure 6.2: Market prices of Stockholm CAT futures (black line), realized CAT values (purple line) and estimated theoretical prices when MPR is zero under different Lévy processes (red for Brownian motion, green for NIG, blue for HYP, light blue for GHYP)

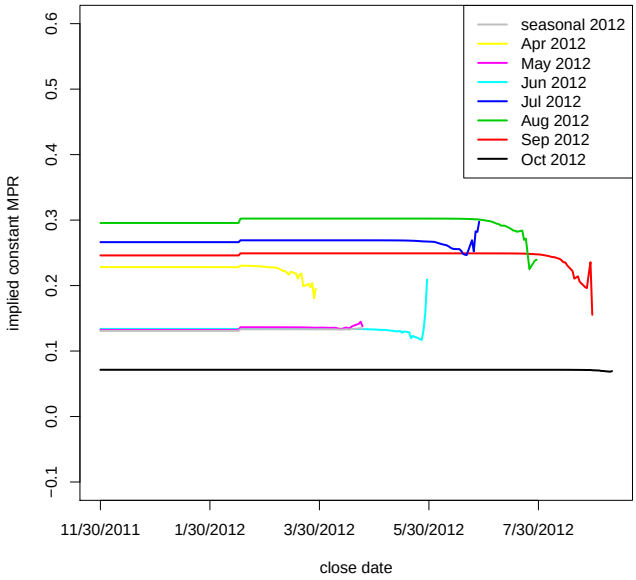



Figure 6.3: Implied Market price of risk for CAT futures

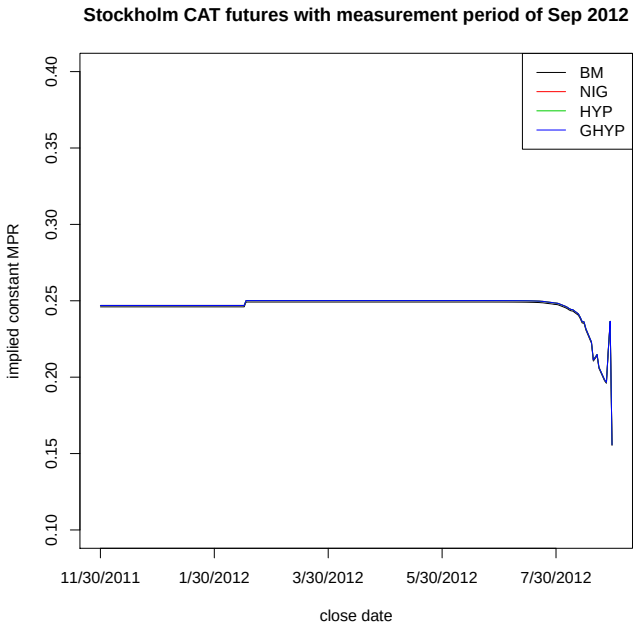



Figure 6.4: Implied Market price of risk for CAT futures with 4 distributions

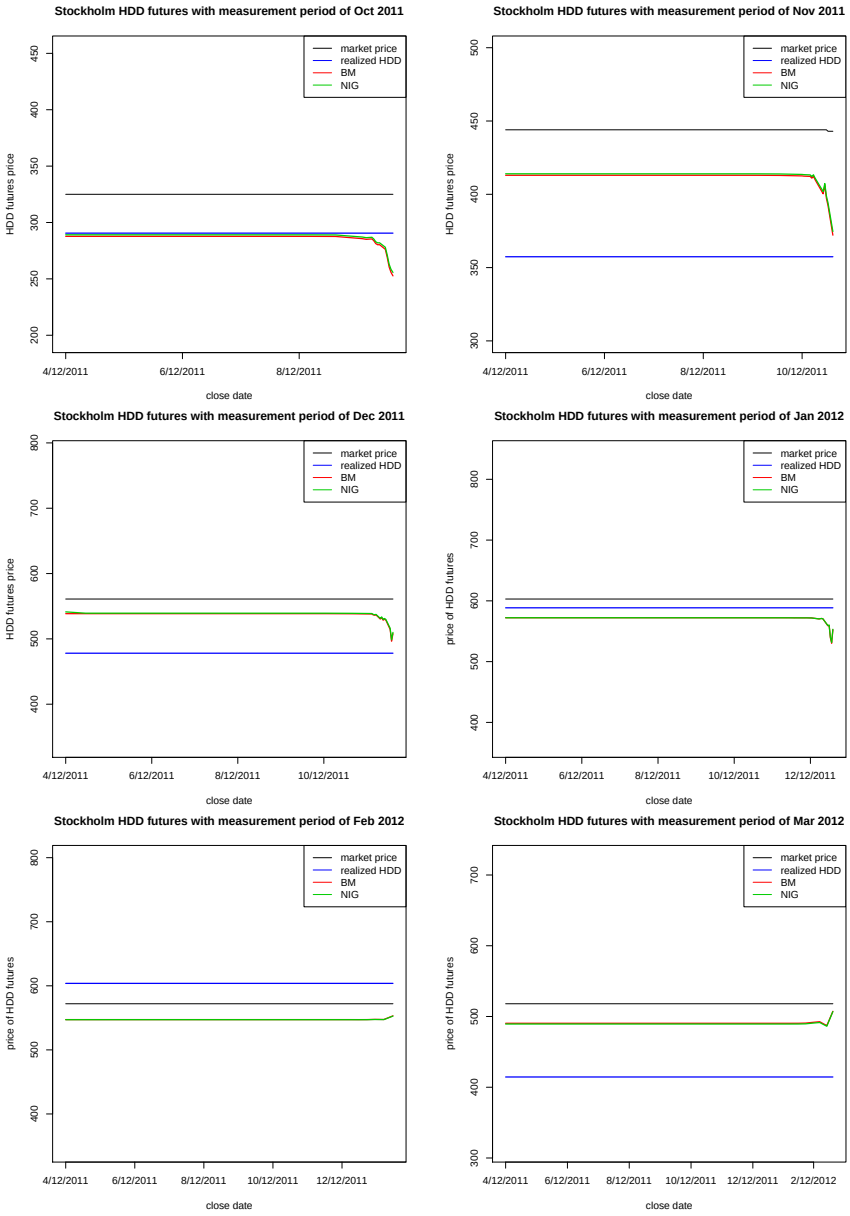



Figure 6.5: HDD Futures Prices of Stockholm with zero MPR

# Chapter 7

# Summary

In this thesis, we explore the framework of pricing temperature derivatives using a Lévy driven continuous-time autoregressive (CAR) model.

Temperature derivatives offered by The Chicago Mercantile Exchange (CME) are introduced. And we focus on futures and options written on three types of temperature indices: the Cumulative Average Temperature (CAT), the Heating Degree Days (HDD), and the Cooling Degree Days (CDD). These temperature indices are based on daily average temperature. Therefore, a stochastic model for modelling the temperature and pricing temperature derivatives is of interest.

We model the daily average temperature using a Lévy driven CAR process with lag $p$ and seasonal volatility. Seasonal variance, autoregressive structure, and mean reverting property of temperature can be explained well by this process. Moreover, we introduce Lévy processes with marginals following the family of generalized hyperbolic distributions. The skewness and heavy tails observed in the standardized residuals can be further captured.

For pricing of temperature derivatives, the arbitrage theory is naturally adopted. Esscher transform is applied to construct a parametrized class of equivalent measures. Under a Brownian motion driven CAR process, we present explicit pricing formulas for the CAT, HDD and CDD futures as well as CAT options. Under the general assumption of a Lévy process, explicit pricing formula for CAT futures is presented. Closed form pricing formulas for CDD and HDD futures and CAT option are not feasible. However, the Esscher transform provides access to the characteristic functions, which enables the use of Fourier technique. Therefore we outline the numerical implementation method here for pricing these contracts.

In the empirical part of this thesis, we analyze the historical temperature data in 6 American cities and 2 European cites. The results show that the family of generalized

hyperbolic distributions provide a better fit than normal distribution. We proceed on implementing the pricing formula as well as the Fourier technique for CAT, HDD and CDD futures and compare it to the realized value and observed market prices. The influence on futures prices from different aspect of the temperature is then discussed.

As a conclusion, We have seen that CAR model is sufficiently good for modelling the temperature evolution and at the same time appropriate for pricing under the arbitrage theory.

For further extension, more sophisticated models or approaches can be used on modelling the seasonal mean, for example wavelet analysis as in Alexandridis and Zapranis [2013]. As we have seen in our data analysis, independence of the residuals is not fulfilled, a generalized autoregressive conditional heteroskedasticity (GARCH) model can be used for calibration of the seasonal volatility, as introduced in Härdle and Cabrera [2012]. For simple CAR(1) process, another interest lies in replacing the constant speed of mean reversion by a time varying one, see for example Alexandridis and Zapranis [2013]. For more sophisticated models using continuous-time autoregressive (CARMA) process and kernel analysis, one can refer to Benth and Benth [2012].

# Bibliography


Hirotogu Akaike. Information theory and an extension of the maximum likelihood principle. In *Breakthroughs in statistics*, pages 610–624. Springer, 1992.

Peter Alaton, Boualem Djehiche, and David Stillberger. On modelling and pricing weather derivatives. *Applied Mathematical Finance*, 9(1):1–20, 2002.

Antonis K Alexandridis and Achilleas D Zapranis. *Weather Derivatives: Modeling and Pricing Weather-Related Risk*. Springer, 2013.

Ted W Anderson and David A Darling. Asymptotic theory of certain" goodness of fit" criteria based on stochastic processes. *The annals of mathematical statistics*, 23(2): 193–212, 1952.

David Applebaum. *Lévy processes and stochastic calculus*. Cambridge university press, 2009.

Ole Barndorff-Nielsen and Christian Halgreen. Infinite divisibility of the hyperbolic and generalized inverse gaussian distributions. *Probability Theory and Related Fields*, 38(4): 309–311, 1977.

Ole E Barndorff-Nielsen. Processes of normal inverse gaussian type. *Finance and stochastics*, 2(1):41–68, 1997.

Francesca Bellini. The weather derivatives market: modelling and pricing temperature. *Institute for Economic Research*, 2005.

Fred Espen Benth and Jūrate Šaltytė Benth. *Modeling and Pricing in Financial Markets for Weather Derivatives*, volume 17. World Scientific Publishing Company, 2012.

Fred Espen Benth and Jūratė Šaltytė-Benth. The normal inverse gaussian distribution and spot price modelling in energy markets. *International journal of theoretical and applied finance*, 7(02):177–192, 2004.

Fred Espen Benth and Jūratė Šaltytė-Benth. Stochastic modelling of temperature variations with a view towards weather derivatives. *Applied Mathematical Finance*, 12(1): 53–85, 2005.

Fred Espen Benth and Jūratė Šaltytė Benth. Weather derivatives and stochastic modelling of temperature. *International Journal of Stochastic Analysis*, 2011, 2011.

Fred Espen Benth, JŪRATĖ ŠALTYTĖ BENTH, and Steen Koekebakker. Putting a price on temperature*. *Scandinavian Journal of Statistics*, 34(4):746–767, 2007.

Fred Espen Benth, Jurate Saltyte Benth, and Steen Koekebakker. *Stochastic modelling of electricity and related markets*, volume 11. World Scientific, 2008.

Fred Espen Benth, Wolfgang Karl Härdle, and Brenda López Cabrera. Pricing of asian temperature risk. In *Statistical Tools for Finance and Insurance*, pages 163–199. Springer, 2011.

Nicholas H Bingham and Rüdiger Kiesel. *Risk-neutral valuation: Pricing and hedging of financial derivatives*. Springer, 2004.

Peter J Brockwell. Lévy–driven continuous–time arma processes. In *Handbook of financial time series*, pages 457–480. Springer, 2009.

Peter J Brockwell and Richard A Davis. *Time series: theory and methods*. Springer, 2009.

PJ Brockwell. Lévy-driven carma processes. *Annals of the Institute of Statistical Mathematics*, 53(1):113–124, 2001.

Dorje C Brody, Joanna Syroka, and Mihail Zervos. Dynamical pricing of weather derivatives. *Quantitative Finance*, 2(3):189–198, 2002.

Rodrigo Caballero, Stephen Jewson, and Anders Brix. Long memory in surface air temperature detection, modeling, and application to weather derivative valuation. *Climate Research*, 21(2):127–140, 2002.

Sean D Campbell and Francis X Diebold. Weather forecasting for weather derivatives. *Journal of the American Statistical Association*, 100(469), 2005.

Kyriakos Chourdakis. Option pricing using the fractional fft. *Journal of Computational Finance*, 8(2):1–18, 2004.

Lesley Choyce. *Caution to the Wind*. Ekstasis Editions, 2000.

Freddy Delbaen and Walter Schachermayer. *The mathematics of arbitrage*, volume 13. Springer Berlin, 2006.

David A Dickey and Wayne A Fuller. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American statistical association*, 74(366a): 427–431, 1979.

Ernst Eberlein. Jump–type lévy processes. In *Handbook of Financial Time Series*, pages 439–455. Springer, 2009.

Ernst Eberlein and Ulrich Keller. Hyperbolic distributions in finance. *Bernoulli*, 1(3): 281–299, 1995.

Fredrik Esscher. On the probability function in the collective theory of risk. *Skandinavisk Aktuarietidskrift*, 14:175–195, 1932.

Hans U Gerber and Elias SW Shiu. Option pricing by esscher transforms. *Transactions of the Society of Actuaries*, 46(99):140, 1994.

Wolfgang Karl Härdle and Brenda López Cabrera. The implied market price of weather risk. *Applied Mathematical Finance*, 19(1):59–95, 2012.

BI Hayman. Maximum likelihood estimation of genetic components of variation. *Biometrics*, 16(3):369–381, 1960.

Sheng-wu He, Chia-kang Wang, and Jia-an Yan. *Semimartingale theory and stochastic calculus*. Taylor & Francis US, 1992.

Edwin Hewitt and Kenneth Allen Ross. *Abstract Harmonic Analysis: Structure and analysis for compact groups, Analysis on locally compact abelian groups. Volume II*, volume 2. Springer, 1963.

Stephen Jewson and Anders Brix. *Weather derivative valuation: the meteorological, statistical, financial and mathematical foundations*. Cambridge University Press, 2005.

Denis Kwiatkowski, Peter CB Phillips, Peter Schmidt, and Yongcheol Shin. Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of econometrics*, 54(1):159–178, 1992.

Greta M Ljung and George EP Box. On a measure of lack of fit in time series models. *Biometrika*, 65(2):297–303, 1978.

Bernt Øksendal. *Stochastic differential equations*. Springer, 2003.

AW Phillips. The estimation of parameters in systems of stochastic differential equations. *Biometrika*, 46(1-2):67–76, 1959.

Karsten Prause. *The generalized hyperbolic model: Estimation, financial derivatives, and risk measures*. PhD thesis, PhD thesis, University of Freiburg, 1999.

Philip Protter. *Stochastic Integration and Differential Equations: Version 2.1*, volume 21. Springer, 2004.

Olivier Roustant, J-P Laurent, Xavier Bay, Laurent Carraro, et al. A bootstrap approach to the pricing of weather derivatives. *Bulletin Français d'Actuariat*, 6(12):163–171, 2004.

David Ruppert. *Transformation and weighting in regression*, volume 30. CRC Press, 1988.

Gideon Schwarz. Estimating the dimension of a model. *The annals of statistics*, 6(2): 461–464, 1978.

Jeremy Staum. Incomplete markets. *Handbooks in Operations Research and Management Science*, 15:511–563, 2007.

A. Swishchuk and K. Cui. Weather derivatives with applications to canadian data. *Journal of Mathematical Finance*, 3(1):81–95, 2013.

Shinzo Watanabe and Nobuyuki Ikeda. *Stochastic differential equations and diffusion processes*. Elsevier, 1981.