UNIVERSITY OF CALGARY

Extreme Value Theory Analysis of Alberta Power Prices

By

Wei Zhang

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE

DEGREE OF MASTER OF ARTS

DEPARTMENT OF ECONOMICS

CALGARY, ALBERTA

APRIL 2005

UNIVERSITY OF CALGARY

FACULTY OF GRADUATE STUDIES

The undersigned certify that they have read, and recommend to the faculty of Graduate Studies for acceptance, a thesis entitled "Extreme Value Theory Analysis of Alberta Power Prices" submitted by Wei Zhang in partial fulfillment of the requirement for the degree of Master of Arts.

W.D. Walls

Supervisor, Dr. David Walls, Department of Economics

Dr. J. R. Boyce, Department of Economics

Robert J. Elliott

Dr. Robert Elliott, Haskayne School of Business

15th April 2005

Date

ABSTRACT

Restructuring of the electricity industry in the past two decades has led to higher volatilities of power prices in competitive wholesale markets. Some electricity market crises, such as the one in California, inspired the use of risk management in newly restructured power markets. This thesis studies high price risks measured by Value at Risk (VaR) in the tails of the underlying distributions of electricity prices. Compared with traditional normal distribution modeling and historical simulation modeling, extreme value theory (EVT) has obvious advantages in estimating VaR by directly dealing with the tails of the underlying distributions and allowing fat-tailed and asymmetric behavior. This argument is confirmed by the empirical data analysis using Alberta Power Pool prices. Back-testing results show that at a high quantile level, VaR estimates from EVT are the most accurate among the above three methods. The estimation results of VaR can be applied in evaluating hedging call options as an insurance against extraordinarily high prices in a wholesale spot power market.

ACKNOWLEDGEMENTS

First, I would like to express my sincere thanks to my supervisor, Dr. David Walls, who introduced me into the risk management analysis area. Without his insights and directions, it would be impossible for me to find the theoretical and analytical tools for my study. I thank him for his encouragement to me that led me to overcome all the difficulties during the past year.

I would also like to thank Dr. Robert Elliott for his valuable suggestions to my thesis from a financial expert's view. I thank him for his insightful discussions with me about my thesis.

I would like to thank Dr. John Boyce and Dr. Frank Atkins during the defense stage of this thesis. Finally, I would like to thank all the other professors in the economics department for their professional instruction and help during my study in the MA program. Special thanks go to Dr. Gordon, Dr. Serletis, Dr. Church, and Dr. Wen for their help with my questions.

DEDICATION

To my mom and dad, Shunzhen and Yucheng, for all their love and support in my life

To my husband, Chunsheng, for his love and support to my venture in Canada

To my son, Weiyi, for his faith in me

TABLE OF CONTENTS

Approval Page....................................ii
Abstract....................................iii
Acknowledgements....................................iv
Dedication....................................v
Table of Contents....................................vi
List of Tables....................................viii
List of Figures....................................ix

Chapter 1 Introduction....................................1

Chapter 2 Extreme Value Theory (EVT) and Risk Management....................................8

2.1 Measures of Extreme Risks: Value at Risk (VaR) and Expected Shortfall (ES)....................................8

2.2 Estimating VaR and ES using Extreme Value Theory (EVT)....................................13

2.2.1 Block Maximum Models and the Generalized Extreme Value Distribution (GEV)....................................14

2.2.2 Peaks-Over-Threshold (POT) Models and the Generalized Pareto Distribution (GPD)....................................21

2.3 Estimating VaR by Other Traditional Approaches....................................26

2.3.1 The Variance-covariance Method with Normal Distribution 26

2.3.2 Historical Simulation Method....................................28

Chapter 3 Estimating Value at Risk (VaR) and Expected Shortfall (ES) Using Extreme Value Theory Models 30

3.1 Data Description....................................30

3.2 Exploratory Data Analysis for Extremes....................................33

3.2.1 Descriptive Statistics for Electricity Prices....................................33

3.2.2 Diagnostic Plots of the Distribution of the Electricity Prices....................................35

3.3 VaR and ES Estimation by EVT....................................41

3.3.1 The Maximum Likelihood Estimation (MLE) Method....................................41

3.3.2 The Maximum Likelihood Estimation of GEV....................................42

3.3.3 The Maximum Likelihood Estimation of GPD....................................47

Chapter 4 Performance of VaR Models..59
4.1 VaR Estimates from Different Models..59
4.2 Back-testing Process and the Comparison of Estimation models............61
4.2.1 Back-testing Process..61
4.2.2 Comparison of the Back-testing Results............................64

Chapter 5 Risk Management Application in the Power Market..........................70
5.1 Electricity Properties and Power Market Structures..........................70
5.2 Application of the GPD Estimated VaR: Hedging Power Price Risks with Call Options..72
5.2.1 Plain-vanilla Options..72
5.2.2 The Application Example: Hedging with Options as Insurance Against Extraordinarily High Power Prices...............................75

Chapter 6 Conclusion..80

References..82

LIST OF TABLES

Table 3-1: Descriptive statistics...35

Table 3-2: Parameter estimates for GEV models...44

Table 3-3: VaR estimates from GEV...45

Table 3- 4: Parameter estimates for GPD models...52

Table 3-5: GPD estimated VaR and ES...54

Table 4-1: VaR estimates from different models...60

Table 4-2: Violation numbers from different models...67

Table 4-3: Violation ratios...69

LIST OF FIGURES

Figure 2-1: Density function of a normal distribution..10

Figure 2-2: Cumulative function of a normal distribution.............................. 11

Figure 3-1: Histogram of winter and non-winter prices..36

Figure 3-2: Exponential QQ-plot of winter and non-winter prices..............................38

Figure 3-3: Sample ME-plot of the winter and non-winter data............................40

Figure3-4: QQ-plot of residuals...46

Figure 3-5: Hill plots..50

Figure 3-6: Plots of GPD fitness...55

Figure 3-7: GPD estimated VaR and ES 57

1. Introduction

The electricity business is comprised of five mutually exclusive services: generation, transmission, ancillary services, distribution, and wholesale and retail supply. Generation is the production of wholesale quantities of power. Transmission is the transportation of wholesale power over large distances using high-voltage cable networks. Ancillary services balance supply and demand in real time and maintain overall system security. Distribution is the transportation of power from the transmission system to consumers. Wholesale and retail supplies are services to facilitate the purchase and sale of the physical commodity of electricity (Masson, G.S., 1999).

Before the 1980s, the electricity industry was an institutionalized natural monopoly and vertical integration was the dominant organizational form of electric utilities. However, the last two decades have seen a trend of restructuring electricity industries around the world. Due to the development of new technologies, more and more jurisdictions have introduced competition in the generation segment and the wholesale (and some of the retail) markets. In the forefront of this restructuring movement is the United Kingdom, where the electricity supply industry (ESI) was privatized in 1990. In Australia the reform of the electricity industry, started in the early 1990s, led to the major re-establishment in southern and eastern Australia of the National Electricity Market (NEM). NEM is a wholesale electricity market commenced on 13 December 1998, including the States of New South Wales, Queensland, Victoria, South Australia and Tasmania participate, together with the Australian Capital Territory, except Western Australia and the Northern Territory (http://www.efa.com.au/reform.html).

In the United States, competition in the electric power industry has been developing for 20 years. There were four key events that increased competition and shaped the new industry (http://www.nei.org/doc.asp?catnum=3&catid=277): the passage in 1978 of the Public Utility Regulatory Policies Act, the emergence through the 1980s of competitive bidding processes to build new power plants, the passage in 1992 of the Energy Policy Act, and the Federal Energy Regulatory Commission's Order 888 in 1996. In Canada, two provincial governments, Alberta and Ontario, have established markets characterized by wholesale and retail unbundling, each with their own specific market designs. The reform to date is considered successful in Alberta but has failed in Ontario. By 2004, more than a dozen countries and jurisdictions have restructured their power industries to include competition, with many more in the planning or early implementation phases.

As Masson (1999) points out, given that providing at least some services of the electricity business does not necessarily require a monopoly market structure, these services can be unbundled and treated as separate markets. Despite differences among jurisdictions, essentially all restructuring of electricity markets separate the generation function from the transmission function, and the distribution function. All the restructuring is designed to provide fair use and equal access to the transmission service. The generation and supply of wholesale power are restructured into an explicitly competitive market, while the transmission and distribution continue to be regulated as natural monopolies. In markets where retail competition exists, the regulated distribution function is separated from the competitive supply function. Ancillary services are a gray area and are treated in various ways, such as a sub-market in some restructures (Masson, G.S., 1999).

The goal of restructuring is to make the electricity industry more efficient and to create benefits for consumers. Efficiency is achieved by introducing price signals to the market players, and by introducing more trading from outside markets. More choice of suppliers and lower prices are a benefit to consumers. Many successes have been achieved in the newly restructured industries, especially in generation segments and wholesale markets. Many challenges remain, one of which is the increased risk in the new industry systems. Risks under a new system include market risk, political risk and trading risk (Doucet, J.A., 2004).

This thesis will discuss market risks under newly restructured electricity industry systems-risks presented by extremely high prices in wholesale spot power markets. As competitively generated electricity becomes a commodity, price volatility is driven by the same supply and demand fundamentals that drive any other commodity market. The unique characteristic in the electricity industry is the high variability in demand and the constrained capacity of supply. Since electricity cannot be stored like other commodities and the transmission systems have finite capacities, electricity markets are necessarily regional in nature. During periods of low demand, the competition among numerous agents at the margin results in low market prices. During periods of high demand and under certain circumstances, only a very few participants are able to supply incremental power. In markets that are temporarily capacity constrained, a few firms may be able to extract substantial profits from extremely high market prices (Masson, G.S., 1999).

A recent example of an electricity market crisis in North America is the one in California in 2000. The California wholesale electricity market was opened in April 1998 when its retail market was still regulated. The wholesale electricity prices increased to extraordinary levels in two years, reaching US$143 per megawatt-hour (MWh) in June 2000. These high prices produced enormous profits for generators and financial crises for the regulated utilities. As Pacific Gas & Electric, the state's largest utility company, declared bankruptcy in March 2001, the state of California had to take over the wholesale electricity purchases and spent more than US$1 billion per month buying power in the spring of 2001. At that time, average prices were more than ten times higher than they had been a year earlier (Borenstein, 2002). The lessons learned from the cases, like the one in California, inspired the use of risk management in the restructured electricity markets.

Starting in finance systems, the risk management revolution has developed enormously in the last two and a half decades, both in theory and in practice (Dowd 1998, 4). In theory, instead of studying average values in the central part of a distribution, risk management analysis only focuses on the extreme values related to the tails of the underlying distribution. This is desirable since many studies suggest that most financial time series have fat-tailed and asymmetric distributions. One definition of fat-tailness (heavy-tailness), given by Gencay and Selcuk (2004), is that a distribution is fat-tailed if a power decay of the density function is observed in the tails. In practice, the development of Value at Risk (VaR) opens up a radically new approach to firm-wide risk management. Value at Risk (VaR) is defined as the maximum expected loss over a given horizon period, at a given level of confidence. The concept of VaR involves two arbitrarily chosen parameters: the horizon period, which might be daily, weekly, monthly, or other time frequencies, and the level of confidence, usually 95% or above. Summarizing the overall market risk faced by a company in a single quantity number, VaR has become a way to quantify market risks. The results of VaR estimates can be used in setting the overall risk targets, allocating assets, assessing the risks of investment, and reporting the levels of risks to investors.

While the underlying distribution is unknown, traditionally the VaR measurement and estimation are based on the assumption of a normal distribution, from which risks of high prices are defined as deviations. By assuming a normal distribution, the deviations from the mean should be symmetric and thin-tailed. Contrary to this assumption, empirical analyses show that financial data are generally asymmetric and fat-tailed.

Extreme value theory (EVT) is the most recent development in tail studies. It deals with the tails of a distribution directly. Unlike the normal distribution model, EVT allows for asymmetric and fat-tailed distributions. The attraction of EVT for risk management analysis is that it fits extreme quantiles very well for fat-tailed and asymmetric distributions of the financial time series. It is a better theoretical foundation for tail studies compared with the traditional methods.

This paper studies modeling and estimating electricity wholesale prices in the upper tails of the underlying price distributions. EVT is used to model the distribution of electricity market prices, and to estimate the probability of extremely high prices. To see the overall advantages of the EVT estimation method, other traditional estimation methods are used as benchmarks. The results of back-testing processes are used to compare the performance of different methods. A back-testing process is a testing framework used in bank systems to test the quality and the accuracy of risk measurement systems. In the process, the actual trading result is compared with the model-generated risk measures. The closer the actual result is to the model-estimated result, the more accurate is the risk measurement model.

The data used in this paper are hourly electricity prices from the Alberta Power Pool from 2001 to 2003. Only extremely high prices in the right side tail are studied as risks (high quantiles at 97%, 99%, and 99.9%) of the underlying distribution, $F(p)$. Similar analysis can be applied to extremely low prices in the left side tail. To do this, simply set the observation data X as negative of the prices P: $X = -P$. The objective of the left tail analysis then becomes an analysis on the right side tail of the negative price distribution, measured by their high quantiles (97%, 99%, 99.9%) of the underlying distribution, $F(x) = F(-p)$.

Using the Alberta data, the comparison of back-testing results indicates that the EVT method is better than other traditional methods, in terms of VaR estimates at high quantile levels. The EVT method of VaR estimation is very useful in the risk management analysis in electricity markets. For example, the EVT estimated VaR can be used to evaluate hedging call options. This thesis demonstrates that EVT is a new reliable tool in risk management analyses in electricity markets.

The arrangement of this thesis is as follows. Chapter 2 introduces extreme value theory (EVT), and applies EVT and other methods in the estimation of Value at Risk (VaR). Chapter 3 first analyzes the empirical data by descriptive statistics and diagnostic plots, and then gives the estimates of VaR from the generalized extreme value distribution (GEV) and the generalized Pareto distribution (GPD). In addition, the estimates of the expected shortfall (ES) are given by GPD. Chapter 4 compares the performance of the three different estimation methods using the back-testing processes. Chapter 5 gives an example of using EVT as a tool for risk management in the restructured power markets. Chapter 6 concludes that EVT is the most reliable estimation tool for analyzing extremely high prices in the restructured Alberta electricity spot market.

2. Extreme Value Theory (EVT) and Risk Management

This chapter introduces measures of risks or extreme values, explains the relationship between extreme value theory and risk estimation, and reviews some traditional estimation methods of VaR.

2.1 Measures of Extreme Risks: Value at Risk (VaR) and Expected Shortfall (ES)

The term Value at Risk (VaR) was first introduced in bank systems. It was quickly adopted among other financial institutions and non-financial corporations. Dowd (1998) and Jorion (1997) thoroughly introduce the origin and development of VaR. Blanco (2001), Dowd (1998), McNeil (1999) and McNeil & Frey (2000) provide the mathematical analyses of VaR and its complementary concept, the expected shortfall (ES).

According to Dowd (1998, 18), in the late 1970s and 1980s, a number of major financial institutions started working on their own internal models to measure and aggregate risks across the institutions as a whole. Among them, JP Morgan developed a best-known system, the RiskMetrics system. This particular system was designed to offer a daily one-page report to JP Morgan's then-chairman, Dennis Weatherstone, indicating risks and potential losses over the next 24 hours, across the bank's entire trading portfolio. The measure used was Value at Risk (VaR) or the maximum likely loss over the next trading day. For the purpose of competing with its rivals and setting an industry standard, JP Morgan published its RiskMetrics system and began to offer the necessary data freely on the Internet in October 1994. This act encouraged many of the software providers to adopt the RiskMetrics approach and to develop VaR software systems. As a result, the use of VaR systems spread quickly among investment banks, commercial banks, pension funds, other financial institutions, and non-financial institutions. In the late 1990s, the VaR methodology was extended further by many users to deal with other risks beside market risks, including credit risks, liquidity risks, cash flow risks, and other risks that are particular concerns for non-financial corporations.

So, what is VaR? According to the Capital Adequacy Directive of the Bank for International Settlement (BIS) in Basle Committee (1996b, 2), VaR is the losses on the bank's trading portfolio over a ten-day holding period in 99% of occasions. For the purpose of internal risk control, most financial firms use a holding period of one day and a confidence level of 95%. From a mathematical point of view, VaR is simply a quantile of the Profit-and-Loss distribution of a given portfolio over a prescribed holding period. The Profit-and-Loss distribution is described by different scenarios of possible future "states" of the market. Each of the VaR estimation methods makes different assumptions about the scenarios, thus has different assumptions of the underlying distributions (Blanco, 2001).

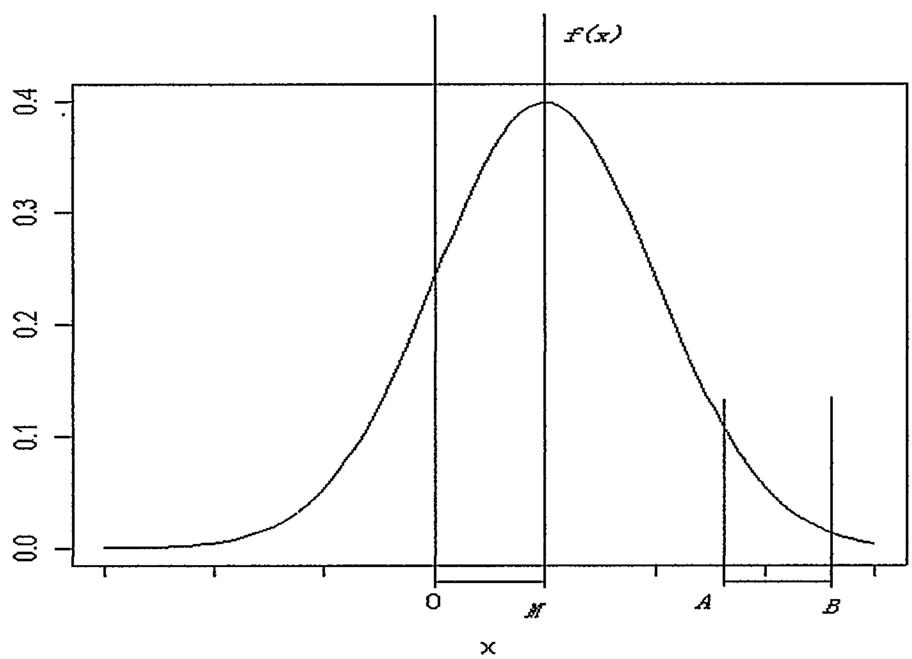
Jorion (1997, 88) defines VaR of a portfolio as W^*, which can be derived from the probability distribution of the future portfolio value $f(w)$. At a given confidence level c, VaR is the worst possible realization W^* such that the probability of exceeding this value is c. The number W^* is a sample quantile of the distribution

$$1 - c = \int_{-\infty}^{W^*} f(w)\, dw = P(w \leq W^*) = p,$$

where $f(w)$ is the density function of the distribution, and $P(w \leq W^*)$ is the cumulative function of the distribution.

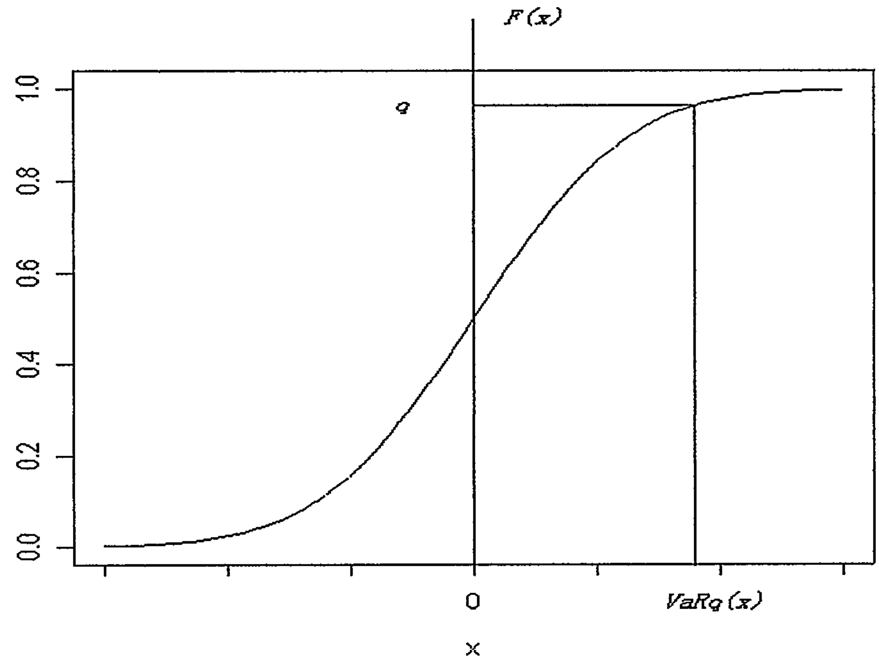
There are two VaR figures, absolute VaR and relative VaR (Dowd, 1998). The former is the maximum amount of the expected loss with a given level of confidence, measured from the current level of value. The latter is measured relative to the mean expected value over the period. Figure 2-1 shows the two types of VaR (under normal distribution assumptions) and their relation.

Figure 2-1 Density function of a normal distribution



Note: μ=mean (= OM); VaR (absolute) = OA; VaR (relative) = OB

Figure 2-2 Cumulative function of a normal distribution



Note: In this case, μ=0, VaR (absolute) = VaR (relative)

In practice, instead of working with two different types of VaR, only the absolute VaR is calculated, since it does not require the measuring of the mean value μ. In fact, in dealing with a short period, the difference between an absolute VaR and a relative VaR is very small.

McNeil (1999) and McNeil & Frey (2000) define more generally the concept of VaR from the statistics point of view. In their description, risks are random variables, mapping unforeseen future states of the world into values representing gains and losses. The potential values of a risk have a probability distribution never observed exactly, although past losses are available. Extreme events, such as losses, occur when a risk takes values from the tail of the underlying distribution. VaR summarizes the measuring of risk with a number, a high quantile of the underlying distribution, typically above

95th percentile. VaR provides a kind of upper bound for a loss, exceeded only in a small proportion of occasions.

Let $X_1, X_2, \ldots X_n$ be identically distributed random variables with unknown underlying distribution function $F(x) = P\{X_i \leq x\}$. Note that there is no assumption of an independent distribution of the random variable here, since the assumption of independence is unnecessary and unrealistic in financial time series (Embrechts et al. 1997).

Measures of extreme risks are defined in terms of the underlying distribution $F(x)$ at high percentiles. Let $0.95 < q \leq 1$, Value at Risk at q (VaR_q) is the qth quantile of the distribution F:

$$VaR_q = F^{-1}(q) \qquad (2\text{-}1)$$

where F^{-1} is the inverse of $F(x)$, and we have $F(VaR_q) = P\{X_i \leq VaR_q\} = q$. See Figure 2-2.

Although VaR gives us a simply summarized number about the loss at a particular point, it tells us nothing about the potential size of the loss that exceeds it. Artzer et al. (1997) propose the use of the expected shortfall (ES) or the tail conditional expectation instead of VaR. Given a high percentile q, the tail conditional expectation is the expected size of a loss that exceeds VaR_q:

$$ES_q = E\,[X \mid X > VaR_q]$$

ES is related to VaR by:

$$ES_q = VaR_q + E\,[X - VaR_q \mid X > VaR_q] \qquad (2\text{-}2)$$

where the second term is simply the mean of the excess distribution $F^{[u]}\ (X-u\,|\,X>u)$ over the threshold VaR_q (McNeil, 1999).

The expected shortfall (ES) is an alternative risk measurement. It provides some information about the size of the potential losses, given that a loss bigger than VaR has occurred. ES is expected to be particularly sensitive with respect to the choice of the model that describes the underlying distribution.

VaR and ES are theoretical quantities that one will never know. The goal in the risk measurement is to estimate them. The main challenge in implementing these risk measures is to come up with a good estimate for the tails of the underlying distribution. Given such tail estimation, both VaR and ES are very easy to compute.

2.2 Estimating VaR and ES using Extreme Value Theory (EVT)

By the nature of a random variable data set, the extreme values relate to the tails of the underlying distribution of the data generating process. The extreme value theory (EVT) is developed as a theory of the study of the tails in a distribution. It explains the general properties of the tails in a distribution, by making the best possible use of the limited set of the realized extreme values. EVT has been widely used in hydrology and structural engineering for studying extreme events (such as earthquakes, floods, etc.). More recently, it became popular in the financial context (Carrillo et al., 2002 and Blanco, 2001).

McNeil surveys all the extreme value theory models in one of his recent papers (McNeil, 1999). There are two main kinds of models for the extreme values. The block maximum models are a group of old models. They are models for the largest observations collected from a large sample of identically distributed observations. The peaks-over-threshold models (POT) are a group of modern models. They are models for all large observations that exceed the high threshold of the underlying distribution.

The block maximum models are based on the generalized extreme value distribution (GEV). While within the peaks-over-threshold models, there are two parametric modeling types. The semi-parametric models of POT are built around the Hill estimators and their relatives; the fully parametric models of POT are based on the generalized Pareto distribution (GPD). With small data sets, the peaks-over-threshold models of EVT will give better estimates of the distribution tails than the block maximum models.

2.2.1 Block Maximum Models and the Generalized Extreme Value Distribution (GEV)

Instead of focusing on the fluctuations of sample averages, in the block maximum models one focuses on the fluctuations of sample block maxima. Sample block maxima, $M_1, \dots M_n$, are defined as

$$M_1 = X_1, \qquad M_n = max\ (X_1, X_2, \dots X_n),\ n \geq 2,$$

where X_1, X_2, $\dots X_n$ are identically distributed random variables with the unknown underlying distribution function $F(x) = P\ \{X_i \leq x\}$ (Embrechts et al, 1997, 115). The block size is represented by n, the number of observations in the block.

Note that only maxima will be studied here, since the corresponding results for minima can be obtained from

$$min\ (X_1, X_2, \ldots X_n) = -\ max\ (-X_1, -X_2, \ldots -X_n).$$

The distribution function of the maximum M_n is

$$P\ (M_n \leq x) = P\ (X_1 \leq x, X_n \leq x) = F^n\ (x), \qquad x \in R,\ n \in N$$

Extremes happen near the upper end of the underlying distribution, thus intuitively the asymptotic behavior of M_n must relate to the underlying distribution function $F(x)$ in its right tail near the right endpoint. If x_F is denoted as the right endpoint of $F(x)$, then

$$P\ (M_n \leq x) = F^n\ (x) \rightarrow 0 \qquad n \rightarrow \infty \quad \text{for all } x < x_F$$

In the case $x_F < \infty$,

$$P\ (M_n \leq x) = F^n\ (x) = 1 \qquad \text{for } x \geq x_F.$$

M_n converges in probability to x_F

$$M_n \xrightarrow{p} x_F, \text{ as n} \rightarrow \infty$$

Since the sequence (M_n) is non-decreasing in n, it converges as sure to the right endpoint of $F(x)$

$$M_n \xrightarrow{a.s} x_F, \text{ as n} \rightarrow \infty.$$

Since the above fact does not provide enough information, one needs to have more insights to the weak convergence results for the centered and normalized maxima, which is one of the major topics in the classical extreme value theory. The first fundamental theorem is the Fisher-Tippett theorem (Fisher & Tippett, 1928). The follow expression of the theorem is from Embrechts et al (1997, 121).

Fisher-Tippett Theorem (Limit Laws for Maxima): Let $X_1, X_2, \ldots$ be a sequence of i.i.d random variables from an unknown distribution F, and $M_n = \max(X_1, X_2, \ldots X_n)$ has the distribution function F^n. If there exist normalizing constants $c_n > 0$ and $d_n \in R$, and some non-degenerate distribution functions H as $n \to \infty$, such that the sequence of the normalized maxima $(M_n - d_n) / c_n$ convergences in distribution to H, then

$$P\{(M_n - d_n) / c_n \leq x\} = F^n (c_n x + d_n) \to H(x), \text{ as } n \to \infty. \qquad (2\text{-}3)$$

The family of the distribution H can be subsumed under a single distribution function, the generalized extreme value distribution (GEV). GEV is the natural limit distribution for the normalized maxima. The definition of the standard GEV is

$$H_\xi(x) = \begin{cases} \exp\{-(1+\xi x)\}^{-1/\xi}, \xi \neq 0 \\ \exp\{-e^{-x}\}, \xi = 0 \end{cases} \qquad (2\text{-}4)$$

where x is such that $1+\xi x > 0$ and ξ is the single parameter. There are three cases of GEV: if $\xi > 0$, there is the Frechet distribution with the shape parameter $\alpha = 1/\xi$; if $\xi < 0$, there is the Weibull distribution with the shape parameter $\alpha = -1/\xi$; if $\xi = 0$, by applying the well-known formula $(1+\xi x)^{1/\xi} \to \exp(x)$ as $\xi \to 0$, there is the Gumbel distribution. Here, the interest is on the Frechet distributions since they are fat-tailed distributions.

Note that if a random variable X has the distribution function $F(x) = P\{X_i \leq x\}$, then $(\mu + \sigma X)$ has the distribution function $F_{\mu,\sigma}(x) = P\{\mu + \sigma X_i \leq x\}$, or $F_{\mu,\sigma}(x) = F((x-\mu)/\sigma)$ (Reiss and Thomas, 2001, 16).

By introducing a location parameter μ and a scale parameter $\sigma > 0$, the family of distributions H can be extended to three-parameter models. The definition of the generalized extreme value distribution $H_{\xi,\mu,\sigma}(x)$ is $H_\xi((x-\mu)/\sigma)$. $H_{\xi,\mu,\sigma}$ is said to be of the type H_ξ as follow:

$$H_{\xi,\mu,\sigma}(x) = H_\xi\left(\frac{x-\mu}{\sigma}\right) = \begin{cases} \exp\left\{-[1+\frac{\xi}{\sigma}(x-\mu)]\right\}^{-\frac{1}{\xi}}, \xi \neq 0 \\ \exp\left\{-e^{-\frac{x-\mu}{\sigma}}\right\}, \xi = 0 \end{cases} . \tag{2-5}$$

Similar to the central limit theorem, where the stable distributions are the only possible non-degenerate limit laws, the Fisher-Tippett theorem says that the family of distribution H is the only possible limiting distribution for the normalized block maxima.

There are two important concepts for furthering the understanding of the Fisher-Tippett theorem. One is the max-stable distribution, and the other is the maximum domain of attraction.

Definition 2-1 Max-stable Distribution A non-degenerate random variable X is called max-stable, if it satisfies

$$F^n(c_n x + d_n) = F(x),\ d_n \in R \tag{2-6}$$

for a suitable choice of constants $c_n > 0$ and $d_n \in R$ and every $n \geq 2$ (Reiss and Thomas, 2001, 19).

Consequently, by combining (2-3) and (2-6), the underlying distribution function $F(x)$ and the standardized maximum function $F^n\ (c_n\ x + d_n)$ are equally distributed, according to the extreme value distribution H, as the number of blocks n tends to infinity.

$$F(x) = F^n\ (c_n\, x + d_n) \rightarrow H(x),\ as\ n \rightarrow \infty, \tag{2-7}$$

Definition 2-2 Maximum Domain of Attraction (MDA) The random variables X_1, $X_2, \ldots$ from an unknown distribution F, are said to belong to the maximum domain of attraction of the extreme value distribution H, if there exist constants $c_n > 0$ and $d_n \in R$ such that (2-3) holds. We write $X \in MDA\ (H)$ or $F \in MDA\ (H)$ (Embrechts et al., 1997, 128).

By definition 2-2, the Fisher-Tippett Theorem says conversely that if F is in the MDA of a non-degenerate extreme value distribution H, then we have the normalizing constants $c_n > 0$ and $d_n \in R$. Reiss and Thomas (Reiss & Thomas, 2001, 19) provide some examples of related constants c_n and d_n, given H follows the Gumble, Frechet, or Weibull distribution.

Gumble:	$d_n = log\ n,$	$c_n = 1,$
Frechet:	$d_n = 0,$	$c_n = n^{1/\alpha},$
Weibull:	$d_n = 0,$	$c_n = n^{1/\alpha}.$

From the Fisher-Tippett theorem and the definition of VaR, one can easily calculate a VaR from its underlying distribution function *F(x)*,

$$VaR_q = H^{-1}(q)$$

where H is derived by the estimated parameters ξ, μ and σ. The explicit expression of the *estimated* H is

$$\hat{\mathrm{H}}(x) = H_{\hat{\xi}}((x-\hat{\mu})/\hat{\sigma})$$

The explicit expression of the *estimated VaR* can be further derived as follows.

From

$$\hat{H}_{\hat{\xi},\hat{\mu},\hat{\sigma}}(x) = \exp\{ -[1 + \frac{\hat{\xi}}{\hat{\sigma}}(x - \hat{\mu})]^{-1/\hat{\xi}}\},$$

there is

$$\ln \hat{H} = -[1 + \frac{\hat{\xi}}{\hat{\sigma}}(x - \hat{\mu})]^{-1/\hat{\xi}}$$

$$1 + \frac{\hat{\xi}}{\hat{\sigma}}(x - \hat{\mu}) = (-\ln \hat{H})^{-\hat{\xi}}$$

$$x - \hat{\mu} = \frac{\hat{\sigma}}{\hat{\xi}}[(-\ln \hat{H})^{-\hat{\xi}} - 1]$$

Substitute $\hat{VaR} = x$ and $q = \hat{H}(x)$ in last equation, there is

$$\hat{VaR} = \hat{\mu} + \frac{\hat{\sigma}}{\hat{\xi}}[(-\ln q)^{-\hat{\xi}} - 1] \qquad (2\text{-}8)$$

Recall the expression of ES_q (2-2):

$$ES_q = VaR_q + E\,[X - VaR_q \mid X > VaR_q]$$

To calculate an ES value, one needs to know the exact excess distribution of the model, which is unknown in GEV. Thus, there is no an explicit equation for the ES estimation from GEV.

The problem of implementing VaR estimation based on GEV is that a large sample of data is required. On the one hand, a sufficiently large block size is required, so that the limiting result of the Fisher-Tippett Theorem may be taken as approximately exact. On the other hand, a large block size also decreases the number of maximum observations drawn from the raw data set. There should be a balance between choosing a large block size, and yet keeping enough maxima data points for the VaR estimation.

2.2.2 Peaks-Over-Threshold Models (POT) and the Generalized Pareto Distribution (GPD)

The generalized Pareto distribution (GPD) is usually expressed as a two-parameter distribution function

$$G_{\xi,\beta}(x) = \begin{cases} 1-(1+\xi x/\beta)^{-1/\xi}, \xi \neq 0 \\ 1-\exp(x/\beta), \xi = 0 \end{cases} \qquad (2\text{-}9)$$

where $\beta > 0$,

$x \geq 0$, if $\xi \geq 0$,

$0 \leq x \leq -\beta/\xi$, if $\xi < 0$.

There are three sub-models for GPD. The case $\xi > 0$ corresponds to the fat-tailed distributions with the tails decaying like the power functions, such as, the ordinary Parteo distribuiton, Student t, Cauchy, Burr, loggamma and Frechet distributions. The case $\xi = 0$ corresponds to the normal, exponential, gamma and lognormal distributions, with the tails decaying exponentially. The case $\xi < 0$ correspond to the distributions with a finite right endpoint, such as the uniform and beta distributions (McNeil & Frey, 2000, 7).

Again, the GPD family can be expended to three-parameter models by adding a location parameter μ. The generalized Pareto distribution $G_{\xi,\mu,\beta}(x)$ is defined to be $G_{\xi,\beta}(x-\mu)$

$$G_{\xi,\mu,\beta}(x) = G_{\xi,\beta}(x-\mu) = \begin{cases} 1-[1+\frac{\xi}{\beta}(x-\mu)]^{-\frac{1}{\xi}}, \xi \neq 0 \\ 1-\exp(-\frac{x-\mu}{\beta}), \xi = 0 \end{cases} \qquad (2\text{-}10)$$

As McNeil points out, the ordinary Pareto distribution is the most relevant distribution for risk management purposes, since GPD is fat-tailed when $\xi > 0$. Whereas a normal distribution has moments of all orders, a fat-tailed distribution does not possess a complete set of moments. In the case of the GPD with $\xi > 0$, the k^{th} moment of the distribution $E[X^k]$ is infinite for $k \geq 1/\xi$. When $\xi = ½$, GPD has an infinite variance. When $\xi = ¼$, GPD has an infinite fourth moment. Mostly in a risk analysis, the assumption is that the underlying distribution F is a distribution with an infinite right endpoint, i.e. it allows the possibility of arbitrarily large losses (McNeil, 1999, 4).

The Generalized Pareto distribution (GPD) functions $G_{\xi,\ \mu,\ \beta}\ (x)$ are the adequate parametric distribution functions for exceedances. The relationship between a GPD function $G\ (x)$ and a GEV function $H\ (x)$ is

$$G\ (x) = 1 + \ln H\ (x),\ if \ln H\ (x) > -1 \qquad (2\text{-}11)$$

(Reiss and Thomas, 2001, 23). It is easy to see by plugging the

$H\ (x)$ expression

$$H_{\xi}(x) = \begin{cases} \exp\{-(1+\xi x)^{-1/\xi}, \xi \neq 0 \\ \exp\{-e^{-x}\}, \xi = 0 \end{cases}$$

into equation (2-11) and compare the result with (2-10).

By (2-11), the same underlying distribution fitted by GEV and GPD models has the same shape parameter ξ, but different scale and location parameters.

The excess (or exceedance) distribution function, or the distribution of excess losses over a high threshold u, is defined as

$$F^{[u]}\ (y) = P\ (X\text{-}u \leq y \mid X > u) \qquad (2\text{-}12)$$

for $0 \leq y < x_F - u$, where $x_F \leq \infty$ is the right endpoint of $F(x)$. The excess distribution $F^{[u]}$ (y) represents the probability that a loss exceeds the threshold u by at most an amount y, given the information that it exceeds the threshold (McNeil, 1999). The exceedance distribution can also be written in terms of the underlying distribution F as

$$F^{[u]}\ (y) = [F\ (y+u)\text{-}F\ (u)]/\ [1\text{-}\ F\ (u)]. \qquad (2\text{-}13)$$

The parametric modeling of the exceedance distribution function $F^{[u]}$ by the generalized Pareto distributions (GPD) is based on a limit theorem, the Pickands-Balkema-de Haan Theorem (Balkema & de Haan 1974, Pickands 1975). The theorem shows that under the MDA condition, GPD is the limiting distribution for the distribution of the excesses, as the threshold tends to the right endpoint. This is a key result in the extreme value theory, which explains the importance of GPD. The follow expression of the theorem is based on MeNeil (1997 and 1999).

Pickands-Balkema-de Haan Theorem For a large class of underlying distributions, one can find a function *β (u)* such that

$$\lim_{u \to x_F} \sup_{0 \le y \le x_F - u} \left| F^{[u]}(y) - G_{\xi,\beta(u)}(y) \right| = 0 . \qquad (2\text{-}14)$$

That is, for a large class of underlying distributions F, as the threshold u is progressively raised, the excess distribution $F^{[u]}$ converges to the generalized Pareto distribution.

In the sense of the Pickands-Balkema-de Haan Theorem, GPD is the natural model for the unknown excess distributions above sufficiently high thresholds. Suppose the right tail of the underlying distribution $F(x)$ begins at the threshold u. Our model for a risk X_i from this underlying distribution $F(x)$ assumes that the excess distribution above the threshold u may be taken as exactly GPD for some ξ and β

$$F^{[u]}(x - u) = G_{\xi, \beta(u)}(x - u). \qquad (2\text{-}15)$$

Again, the choice of a threshold is a compromise between choosing a sufficiently high threshold, which ensures the asymptotic exact of the Pickands-Balkema-de Haan Theorem, and choosing a sufficiently low threshold, which gives enough excedances for the estimation of the GPD parameters (McNeil, 1999).

From the result of the Pickands-Balkema-de Haan Theorem, it is easy to calculate the tail estimates of the underlying distribution F. By setting x = u + y, one can rewrite the excesses distribution function (2-13) as

$$F(x) = [1 - F(u)]F^{[u]}(y) + F(u). \qquad (2\text{-}16)$$

From (2-14), we have $F^{[u]}(y) \rightarrow G_{\xi,\beta(u)}(x-u)$ for large u, where x > u. Equation (2-16) can be rewritten as

$$F(x) = [1-F(u)]\, G_{\xi,\beta(u)}(x-u) + F(u), \qquad \text{for } x > u.$$

This formula interprets the model in terms of the tail estimation of the underlying distribution $F(x)$, given $x > u$ and the estimate of $F(u)$. One suggestion for estimating $F(u)$ from McNeil (1999) is to use the historical simulation (HS) estimator $F(u) = \frac{n-n_u}{n}$. Here n is the sample size and n_u is the number of exceedances given a threshold u.

Putting the HS estimate of $F(u)$ and the maximum likelihood estimates (MLE) of the parameters of GPD together, one can derive the tail estimator as follows.

$$\hat{F}(x) = (1 - \frac{n-n_u}{n})G_{\hat{\xi},\hat{\beta},u}(x-u) + \frac{n-n_u}{n}$$

$$= \frac{n_u}{n}G_{\hat{\xi},\hat{\beta},u}(x-u) + \frac{n-n_u}{n}$$

$$= 1 + \frac{n_u}{n}(G_{\hat{\xi},\hat{\beta},u}(x-u) - 1)$$

Given $G_{\xi,\mu,\beta}(x) = 1 - [1 + \xi(\frac{x-\mu}{\beta})]^{-\frac{1}{\xi}}$, plug in the equation above, one has

$$\hat{F}(x) = 1 - \frac{n_u}{n}(1 + \hat{\xi}\frac{x-u}{\hat{\beta}})^{-\frac{1}{\hat{\xi}}}$$

$$\frac{n_u}{n}(1 + \hat{\xi}\frac{x-u}{\hat{\beta}})^{-\frac{1}{\hat{\xi}}} = 1 - \hat{F}(x) = 1 - q$$

where, q is the high probability above F (u). Let $\alpha = 1 - q$, one has

$$(1 + \hat{\xi}\frac{x-u}{\hat{\beta}})^{-\frac{1}{\hat{\xi}}} = \frac{n}{n_u}\alpha$$

$$\hat{\xi}\frac{x-u}{\hat{\beta}} = \left[\frac{n}{n_u}\alpha\right]^{-\hat{\xi}} - 1$$

For a given probability $q > F(u)$, the VaR estimate x is calculated by

$$x = \hat{VaR}_q(1-q) = u + \frac{\hat{\beta}}{\hat{\xi}}[\frac{n}{n_u}(1-q)^{-\hat{\xi}} - 1]. \qquad (2\text{-}17)$$

The GPD model for the excess distribution above a threshold u has a nice stability property. If one takes any higher threshold above u, such as $u' = VaR_q > u$, for $q > F(u')$, the excess distribution $F^{[u']}(y)$ above the higher threshold u' is also a generalized Pareto distribution, with the same shape parameter ξ, but a different scale parameter β (McNeil, 1999). A consequence of the model $F^{[u]}(y) = G_{\xi,\beta(u)}(y)$ is that

$$F^{[u']}(x - u') = G_{\xi,\beta(u')}(x - u')$$

$$F^{[u']}(x - VaR_q) = G_{\xi,\beta(u')}(x - VaR_q). \qquad (2\text{-}18)$$

Equation (2-18) is a simple explicit expression for the excess losses above VaR_q. With this expression, one can calculate many losses beyond VaR_q. Noting that (providing $\xi < 1$) the mean of the distribution in (2-18) is

$$E\,[x - VaR_q] = (\beta + \xi\,(VaR_q - u))/(1 - \xi), \tag{2-19}$$

one can calculate the expected shortfall (ES) by plugging (2-19) into (2-2). The result is

$$ES_q = VaR_q + \frac{1}{1-\xi}[\beta + \xi(VaR_q - u)]$$

$$ES_q = \frac{VaR_q}{1-\xi} + \frac{\beta - \xi u}{1-\xi} \tag{2-20}$$

The estimated ES is expressed by substituting the data based estimates of what is unknown in (2-20) to obtain

$$\hat{ES}_q = \frac{\hat{VaR}_q}{1-\hat{\xi}} + \frac{\hat{\beta} - \hat{\xi} u}{1-\hat{\xi}} \tag{2-21}$$

2.3 Estimating VaR by Other Traditional Approaches

Several traditional backward-looking methods have been used in measuring VaR. Two of them are introduced in this chapter, the variance-covariance method with the normal distribution assumption and the historical simulation method with no distribution assumption.

2.3.1 The Variance-Covariance Method with Normal Distribution

The variance-covariance approach of VaR has the same theoretical basis as the portfolio theory. They all interpret risks in terms of the standard deviation of the return. The variance-covariance approach is the simplest approach among the various models for VaR estimations. A most straightforward assumption of the underlying distribution is the normal distribution assumption (Dowd, 1998, 42).

Suppose X is identically distributed random variable with an underlying normal distribution function $N(x) = P\{X_i \leq x\}$. One can always describe the confidence level in terms of a single parameter, α, which tells how far away the cut-off values of the two tails are from the mean, μ, in terms of units of the standard deviation, σ. In studying extremely high values above X* in the right tail of the distribution, one consider the probability function

$$Prob\{X \geq X^*\} = Prob\{Z \geq (X^* - \mu)/\sigma\} = 1 - N(X^*) = 1 - q = c,$$

where Z is the standard normal variate with $\mu = 0$ and $\sigma = 1$, and c represents the probability of a right-tail event (such as $q = 1 - c = 95\%$ confidence level).

In general, one has

$$X^* = \mu + \alpha\sigma$$

where α reflects the selected confidence level. From the standard normal tables, one can read off the value of $\alpha = (X^* - \mu)/\sigma$. For example, with $c = 95\%$ confidence level, there is $\alpha = (X^* - \mu)/\sigma = 1.65$.

VaR_q, with the assumption of a normal distribution, is expressed as

$$VaR_q = N^1(q)\,\sigma \tag{2-22}$$

where $N^{1}\,(q)$ is the qth quantile value of the normal distribution. The estimate of μ and σ can be obtained from the sample mean and the sample variance by

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{2-23}$$

$$\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \hat{\mu})^2 \tag{2-24}$$

The normal distribution assumption in the variance-covariance method is simple to use. The problem is that it does not account for asymmetric possibilities. In addition, a normal distribution has an exponential decay at its tails. As a result, the normal distribution estimation tends to underestimate the risks at high quantiles in the right tail (or at low quantiles in the left tail), given the fact that empirical distributions often exhibit asymmetric fat-tails.

2.3.2. Historical Simulation Method

Based on Carrial et al. (2002) and Dowd (1998), the basic assumption of the historical simulation is that the underlying distribution is unknown, and that future risks are much like past risks. Thus, the VaR estimates from the historical simulation method are high quantiles of the empirical distribution. These high quantiles are obtained from the historical data of the realized observations.

Among all the methods of modeling risks, historical simulations do not rely on assumptions about the underlying stochastic structure of the distribution. The advantage of a historical simulation is that it accounts for the fat-tailness and does not suffer from model specific errors. In this sense, it is better than the variance-covariance method with the normal distribution assumption in describing risk scenarios.

A disadvantage of the historical simulation is the assumption that future risks are much like past risks. When there are few observations in the historical observation sample, some extreme values not found in the past will not be predicted in the future. As such, the historical simulation is a poor method in analyzing the tails of the underlying distribution, where the sample data becomes sparse.

3. Estimating Value at Risk (VaR) and Expected Shortfall (ES) Using Extreme Value Theory Models

3.1 Data Description

The raw data of hourly electricity prices in this paper are Alberta Power Pool prices from January 2001 to December 2003. The data is obtained from the website of Alberta Electric System Operator, www.aeso.ca. A whole set of historical data is available from 1999 to date. The reason to choose the data beginning from January 2001 is that this was when Albertans could choose their electricity suppliers and thus retailers started real market competition.

The Alberta Power Pool began to operate in 1996, when the Electric Utilities Act came to effect. The pool is the wholesale electricity spot market where all power in Alberta must be bought and sold. The participants are suppliers and consumers. The suppliers include both power producers, and companies that purchased the rights, during two auctions in 2000, to the formerly regulated Alberta generation. The consumers include retailers who purchase the power and provide retail services for many other customers, and self-retailers who buy the power for their own use. By now, there are more than 200 participants and about \$ 3 to \$5 billion in annual energy transactions in the spot market (AESO fast facts, www.aeso.ca).

The heart of the Alberta real-time electricity market is the System Coordination Center (SCC), which is staffed 24 hours a day, seven days a week by a team of system controllers. System controllers dispatch electricity to meet real-time demand, which sets the hourly real-time wholesale market prices. The prices are established in a "merit order dispatching system": First, power producers and importers submit supply offers, and exporters and consumers submit demand bids. The supply offers and demand bids are sorted from the lowest price to the highest price at each hour of the day. Then, as electricity demand shifts throughout the day, system controllers keep supply and demand in balance by dispatching the next supply offers or demand bids in the merit order. Every minute, the last eligible electricity block dispatched by the system controller sets the System Marginal Price (SMP), which is updated in real-time and published on the AESO website. At the end of each hour, the time-weighted average of the sixty one-minute SMP is calculated and published as the market price.

Because of the above open and transparent operation, the real time electricity wholesale market, the Alberta Power Pool, demonstrates similar characteristics as stock markets. The price data from this electricity wholesale market has properties similar to other time series.

The differences between this electricity spot market and stock markets come from the special properties of electricity supply and demand. From the demand side, the electricity demand is subject to cyclical, seasonal, and daily fluctuations. In the long run, demand increases when the economy goes up and decreases when the economy goes down. At a seasonal level, demand is different between winter and summer. At a daily level, demand is much higher in peak hours than in off-peak hours. From the supply side, supply is constrained by the capacity of generators. Meanwhile, the diversity of generation technologies and cost structures result in various supply prices. A coal-fueled generator cannot stop its' generation process during off-peak hours, because it needs much longer time to preheat for a newly started operation. A quickly started gas-fueled generator requires high prices to cover its high fuel cost. Each day during off-peak hours, when many generators keep offering more supply of power than the demand of it, the price stays in a very low range. In contrast, when demand increases sharply in peak hours and supply is constraint by its capacity, the price goes up quickly. As a result, the real spot electricity wholesale market displays frequent fluctuations in its' prices.

A study of the price distribution needs filtering the raw data, given the special stochastic properties of the power pool prices. To find an identically distributed time series, one can filter the data of electricity prices by seasons and peak hours. The assumption here is that the seasonal and peak-hour factors count for most non-stationary sources and that filtering the raw data by seasons and peak-hours gives the identically distributed observations.

The convenience of Alberta Power Pool prices is that they are recorded by year, month, day, and hour. Thus, one can filter hourly prices according to the choices of different seasons and peak hours. According to weather conditions in Alberta, it is reasonable to define a two-season pattern in a year for the analysis: the eight- month winter season from October to May, and the four-month non-winter season from June to September. To decide the accurate peak hours, each average hourly price is calculated from the whole set of raw data. Among the twenty-four average hourly_prices, ten of them from 11:00 am to 20:00 pm are considered as peak hour prices, ranging from CD$71.57 per MWh to CD$92.82 per MWh.

To get the identically distributed observations, the set of raw data is filtered by the peak hours and the two seasons. First, by filtering out fourteen off-peak hour prices, there are then ten peak hour prices for the daily-based prices. Then, the whole set of peak hour prices are divided into two seasons: the winter season data from October to May, and the non-winter data from June to September. After filtering, there are two sets of data for further analysis, one is the winter-season peak-hour data with 7290 observations, and the other is the set of non-winter-season peak-hour data with 3660 observations.

3.2 Exploratory Data Analysis for Extremes

It is always useful to look at the data before carrying out a detailed statistical analysis. Two kinds of tools are used in this paper for the exploratory data analysis, one is the descriptive statistics, and the other is the diagnostic plot.

3.2.1 Descriptive Statistics of the Electricity Prices

A set of basic descriptive statistics of a distribution includes the minimum observed value, the maximum observed value, the mean, the standard deviation, the Skewness, and the Kurtosis. The follow descriptions of these concepts are based on Pindyck & Rubinfeld (1998).

Suppose there is a sequence of discrete variables $X_1, \ldots, X_N$ that represent N possible outcomes associated with the random variable X. The expected value, or the mean, of X is a weighted *average* of the possible outcomes, where the probabilities of the outcomes serve as the appropriate weights. The average is denoted $\mu = E(X)$.

The variance of a random variable provides a measure of the spread, or *dispersion*, around the mean. The variance is denoted σ^2, and defined as $Var\ (X) = \sigma^2 = E[(X-\mu)^2]$.

Skewness is a statistic that provides useful information about the *symmetry* of a probability distribution. The skewness statistic S of a variable X is given by

$$S = (1/N) \sum (X-\mu)^3 / \sigma^3,$$

where S is equal to zero for all symmetric distributions. For a non-symmetric distribution, S is positive when the upper tail of the distribution is thicker than the lower tail, and is negative when the lower tail is thicker than the upper tail.

Kurtosis provides a measurement of the *thickness*, or the shape of the tails of a distribution. The kurtosis statistic, K, is given by

$$K = (1/N) \sum (X-\mu)^4 / \sigma^4,$$

where for a normal distribution, K is equal to three. When the tails of the distribution are thicker than a normal distribution, we have $K > 3$, and vice versa.

The descriptive statistics of Alberta Power Pool prices are shown in Table 3-1. It is clear that the two sets of data both disperse largely from their means. Their skewness statistics are both positive, which means that the underlying distributions are asymmetric with the upper tails thicker than the lower tails. Their kurtosis statistics are both much larger than *3,* which mean that the tails of the underlying distribution are thicker than that of normal distributions. In other word, they are fat tails.

Table 3-1 Descriptive statistics

Data	n	μ	σ	S	K	Min	Max
Winter	7290	80.62	79.54925	4.264335	30.1544	9.81	1000
Non-winter	3660	67.67	78.05554	6.382265	59.97808	6.29	999

Note: Winter = October to May from 2001 to 2003 peak hourly prices (from 11:00 am to 20:00 pm); Non-winter = June to September from 2001 to 2003 peak hourly prices (from 11:00 am to 20:00 pm); n = sample size; μ = sample mean; σ = sample standard deviation; S = Skewness; K = Kurtosis; Min = Minimum observed price; Max = Maximum observed price

3.2.2 Diagnostic Plots of the Distribution of the Electricity Prices

With the descriptive statistics above, one gets some basic ideas about the shape of the distributions of Alberta Power Pool prices. Diagnostic plots can describe further details more clearly. For studying the tail behavior, the histogram, the QQ-plot, and ME-plot will be used in this paper. The follow description of the plot concepts is based on McNeil (1997), Reiss & Thomas (2001), and Embrechts et al (1997). The graphs are generated by the R-project software (http://www.r-project.org).

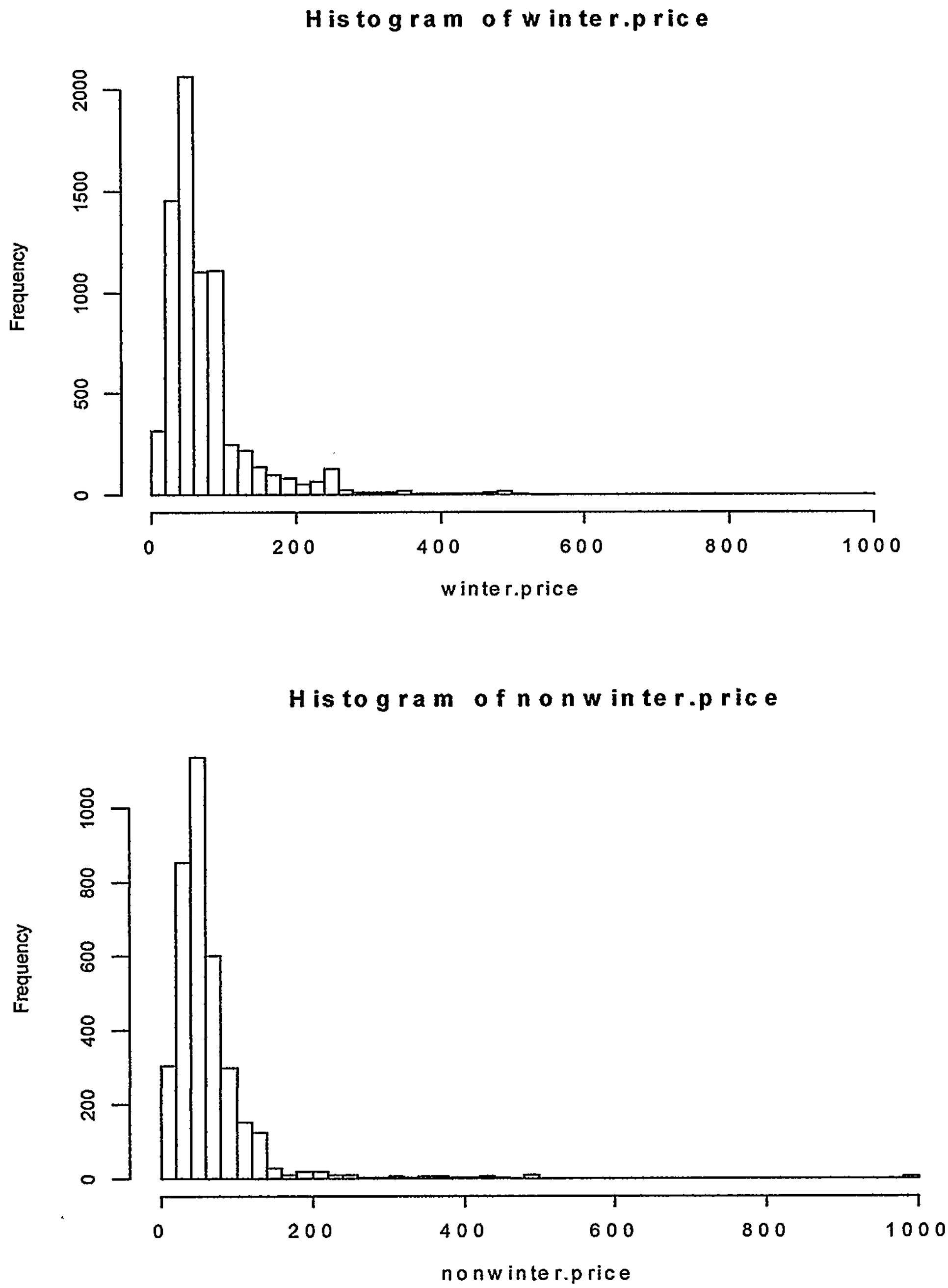


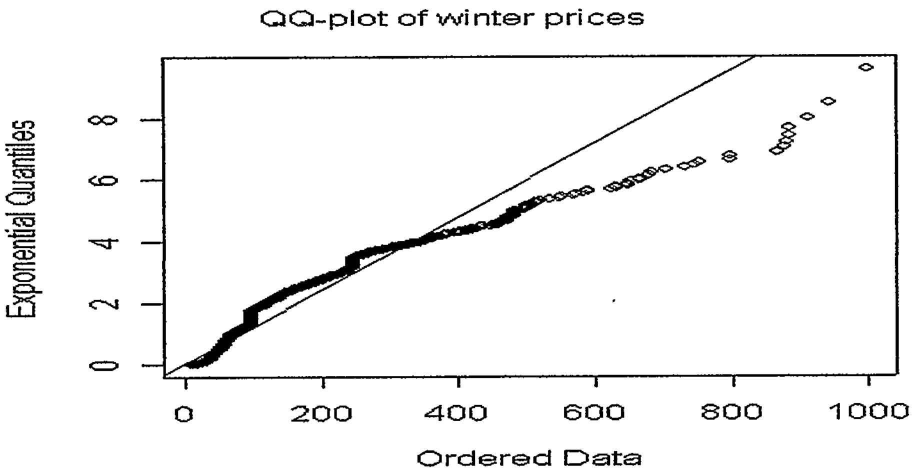
Figure 3-1 Histogram of winter and non-winter prices

Given a set of data, one usually starts diagnostic plot analysis with a histogram. A histogram is usually plotted of the sample observation values against their frequency numbers. The histograms of the Albert Power Pool data are shown in Figure 3-1.

A quantile plot, QQ-plot, is the plot of the sample quantiles against the quantiles of the assumed underlying (or tail) distribution. An assumed tail distribution can be any one of the generalized Pareto distributions (GPD), such as the exponential, the student t, or the beta distribution. Since an underlying distribution has the same shape parameter with its related tail distribution, an exponential tail distribution has an underlying Gumbel distribution. To find the evidence of a fat tailed distribution, an exponential distribution is used as the assumed distribution. In a QQ-plot, the quantiles of the empirical distribution is plotted on the x-axis, with the quantiles of the exponential distribution on the y-axis. That is

$$\left\{(X_{k:n}, G_{0,1}^{-1}(\frac{n-k+1}{n+1})), k=1,\ldots,n\right\},$$

where $X_{k:n}$ denotes the k^{th} order statistic, and $G_{0,1}^{-1}$ is the inverse of the exponential distribution function. If the points lie approximately along a straight line, the set of data is a sample from an exponential distribution. A concave shape of the points displays a fat tailed distribution, whereas a convex shape shows a thin-tailed distribution.



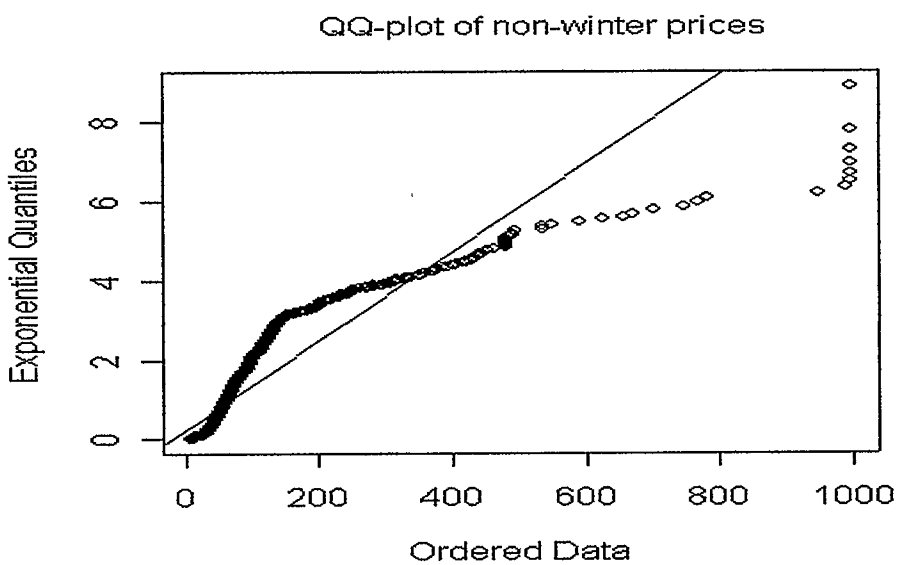


Figure 3-2 Exponential QQ-plot of winter and non-winter prices

The QQ-plots of the Alberta data are shown in Figure 3-2, where the empirical quantiles of electricity prices are plotted against the exponential quantiles. From the concave shape of the plots, one can conclude that the two sets of electricity prices are from fat tailed distributions, with their shape parameter $\xi > 0$.

Another useful graphical tool, in particular for distinguishing different tails, is the sample mean excess plot. A sample mean excess plot (ME-plot) is the plot of the sample mean excess function $\{(u,\ e_n(u)),\ X_{n,n} < u < X_{1,n}\}$, where $X_{1,n}$ and $X_{n,n}$ are the 1^{st} and n^{th} order statistics and $e_n(u)$ is the sample mean excess function.

A sample mean excess function (sample ME) is an empirical estimate of the mean excess function, which is expressed as

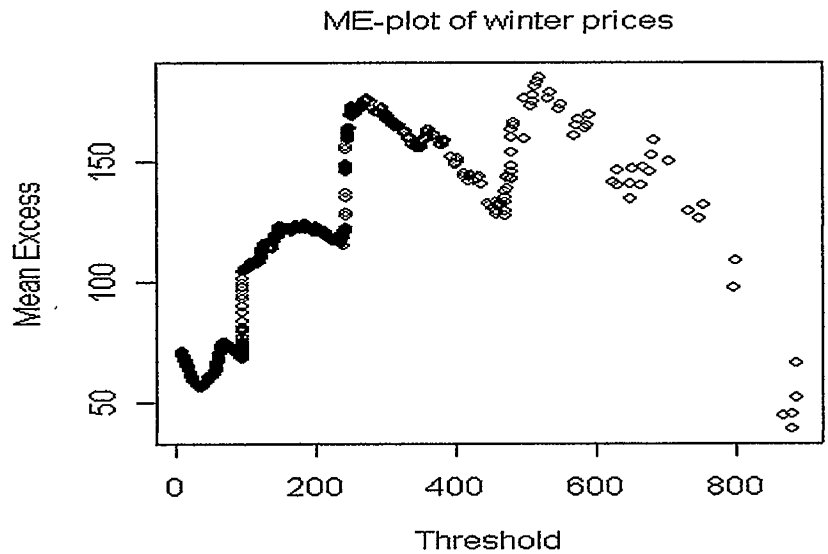
$$e_n(u) = \frac{\sum_{i=1}^{n}(x_i - u)^+}{\sum_{i=1}^{n} I_{\{x_i > u\}}}$$

where I is an indicator function equal to one when $x_i > u$, and equal to zero when $x_i \le u$. A mean excess function (ME) is the sum of the excesses over the threshold u divided by the number of data that exceed u, which is expressed as

$$e\,(u) = E\,(X - u \mid X > u),\ 0 \le u < x_F,$$

where x_F is the right endpoint of the underlying distribution.

The ME-plot is used mainly for distinguishing thin-tailed distributions from fat-tailed distributions. If the points in a sample ME-plot show an upward trend, then it is considered an evidence of a fat-tailed distribution. An approximately horizontal line of the points relates to an exponential distribution, whereas a downward trend of points relates to a thin-tailed distribution. In particular, if the empirical plot follows a straight line with a positive slope above a certain value of u, then the data follow a generalized Pareto distribution (GPD) in the tail above the threshold u (McNeil, 1997).



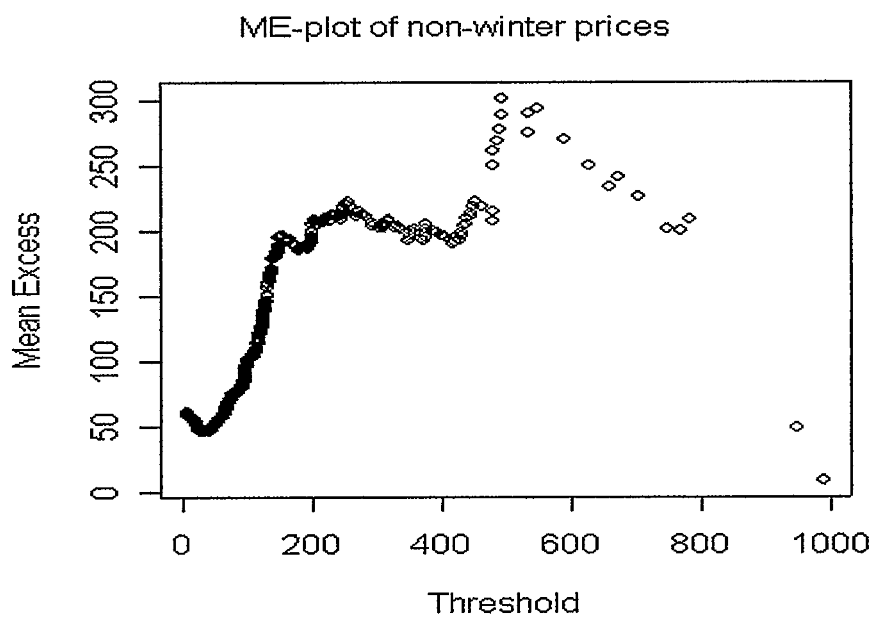


Figure 3-3 Sample ME-plot of winter and non-winter data

From Figure 3-3, the points in the sample ME-plots of the Alberta data are upward trending, which means that the electricity prices of the Alberta Power Pool are from fat-tailed underlying distributions.

All the above descriptive statistics and diagnostic plots show that the two data sets of the Alberta Power Pool prices are realized observations from asymmetric fat-tailed underlying distributions.

3.3 Estimating VaR and ES by EVT

3.3.1 Maximum Likelihood Estimation (MLE) Method

The generalized extreme value distribution (GEV) and the generalized Pareto distribution (GPD) are both parameter distributions. To estimate VaR and ES with GEV and GPD models, first one needs to estimate the related parameters. The maximum likelihood estimation (MLE) method is the most often used method for parameter estimations.

When using MLE to estimate the parameters in GEV, the assumption is that the data $X_1, \ldots X_n$ consists of a sample of identically independent distribution of $H_{\xi, \sigma,\mu}$, which is $X_1, \ldots X_n$ i.i.d from $H_{\xi, \sigma, \mu}$. In fact, the above condition is hard to be satisfied. It can be shown that relaxing the condition to

$$X_1, \ldots X_n \text{ i.i.d from } F \in MDA(H_\xi, \sigma, \mu),$$

the MLE method is still available (Embrechts et al., 1997, 316).

The maximum likelihood estimation is the most general fitting method in statistics. It allows one to estimate the statistical errors (standard errors) for the parameter estimates.

After estimating the parameters by MLE, one can calculate the estimated VaR from the GEV and GPD models and ES from the GPD model. It is possible to give a confidence interval for the estimated VaR and ES in the GPD model, using a method known as profile likelihood. The confidence intervals of the estimated VaR and ES are asymmetric, which reflect fundamental asymmetries in the heavy-tailed distributions.

3.3.2 Maximum Likelihood Estimation of GEV

The key issue of fitting GEV for empirical data is to decide a block size n to get the maxima for estimating the parameters. Recall from the limit law for maxima (the Fisher-Tippett Theorem), condition (2-3) holds only if the block size $n \to \infty$. Consequently, the larger the block size n, the better the fitness of the GEV for the data. The problem is that given a set of raw data, the larger the block size n, the less the resulted observed maxima. If one chooses too large a block size n, one will have sparse data for the estimation and consequently get large standard deviations from the estimated parameters.

In this analysis, since the raw data of prices are ten hourly prices of peak-hours, the maxima block sizes are chosen from 10, 30, 50, 70, 80, 90, to 110, which are related to one day, three days, ...to eleven days of peak hours data. The estimation results of the parameters and the VaR are listed in Table 3-2 and Table 3-3.

From Table 3-2, one can see that the estimated value of the shape parameter ξ decreases as the block size increases. From the Fisher-Tippett Theorem, we say that the estimates from large size blocks should be considered correct. As a result, it seems that the estimates from small size blocks tend to exaggerate a fat-tailedness fact and give much larger estimates of ξ values. A small increase in the block size results in much fewer observations in the maxima data sets. According to the asymptotic distribution theory, the estimations from small samples are less accurate than from large samples. This fact is confirmed by the large standard deviations of our parameter estimates from small samples.

Table 3-2 Parameter estimates for GEV models

Raw data	Block size	Maxima	Estimated parameters & standard deviations					
N	n	k	ξ	s.e. (ξ)	σ	s.e.(σ)	μ	s.e.(μ)
	10	729	0.571205	0.034617	52.3642	2.22079	77.5455	2.16829
Winter	30	243	0.587092	0.093875	90.0274	7.2798	121.921	7.32328
N=7290	50	144	0.423161	0.112003	121.871	11.5224	164.66	12.6746
	70	105	0.262945	0.132234	153.564	16.1165	210.788	19.1002
	80	92	0.178212	0.131228	167.989	17.8041	236.346	21.8923
	90	81	0.121395	0.124853	178.495	18.9851	254.244	24.0494
	10	366	0.526774	0.046607	43.6281	2.50154	60.9599	2.53625
Non-winter	30	122	0.599027	0.096425	78.9698	8.4532	98.1616	8.19481
N=3660	50	74	0.595754	0.157017	102.94	14.8089	133.1727	14.7122
	70	53	0.526208	0.196303	130.142	21.7604	168.7567	22.6411
	80	46	0.462364	0.187255	136.339	22.9293	184.571	24.6125
	110	33	0.205696	0.203841	190.207	32.6987	247.574	40.0149

Substituting the estimated parameters into the VaR estimation equation (2-8)

$$VaR = \mu + \frac{\sigma}{\xi}[(-\ln p)^{-\xi} - 1].$$

The calculated VaR estimates are shown in Table3-3. From Table 3-3, one can see that when the block size n > 10, a larger estimated shape parameter ξ always relates to a higher estimated VaR.

Table 3-3 VaR estimates from GEV

Raw data	Block size	Maxima	Estimated	VaR (q) estimates		
	n	k	ξ	q = 0.97	q = 0.99	q = 0.999
	10	729	0.571205	634.21	1229.53	4700.19
Winter	30	243	0.587092	1127.56	2220.04	8784.13
N=7290	50	144	0.423161	1095.77	1851.27	5189.18
	70	105	0.262945	1032.14	1527.22	3160.44
	80	92	0.178212	981.508	1365.18	2453.4
	90	81	0.121395	954.564	1278.23	2108.94
	10	366	0.526774	481.857	895.226	3111.09
Non-winter	30	122	0.599027	1041.49	2020.92	8206.89
N=3660	50	74	0.595754	1313.24	2607.64	10514.18
	70	53	0.526208	1435.71	2665.93	9253.52
	80	46	0.462364	1323	2315.23	7029.84
	110	33	0.205696	1161.75	1647.52	3094.14

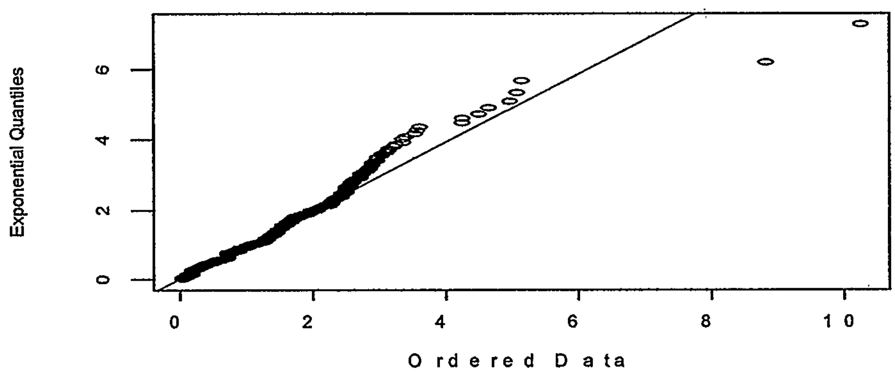
After estimating parameters by choosing different block sizes, one can check the fitness of GEV models to the data by residual plots, where the residual is defined as

$$res = \begin{cases} \left[1 + \frac{\xi}{\sigma}(x - \mu)\right]^{-\frac{1}{\xi}}, \xi = 0 \\ \exp\left[-\exp\left(-\frac{x - \mu}{\sigma}\right)\right], \xi \neq 0 \end{cases}$$

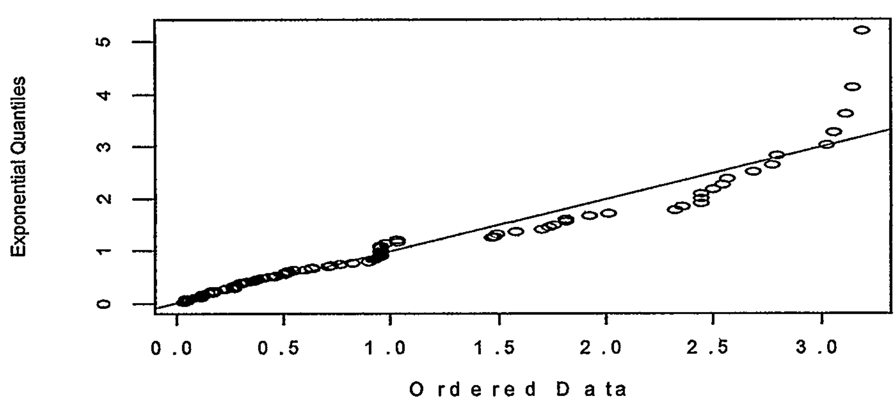
The observations are converted to exponentially distributed residuals under the null hypothesis that GEV fits the data. Figure 3-4 shows QQ-plots of residuals with different block sizes for the Albert data.

Figure3-4 QQ-plot of residuals

(a) Winter data

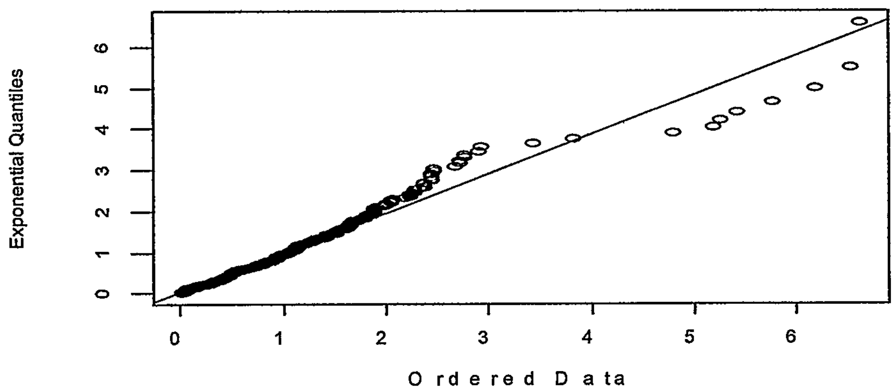


n = 10 (k = 729)

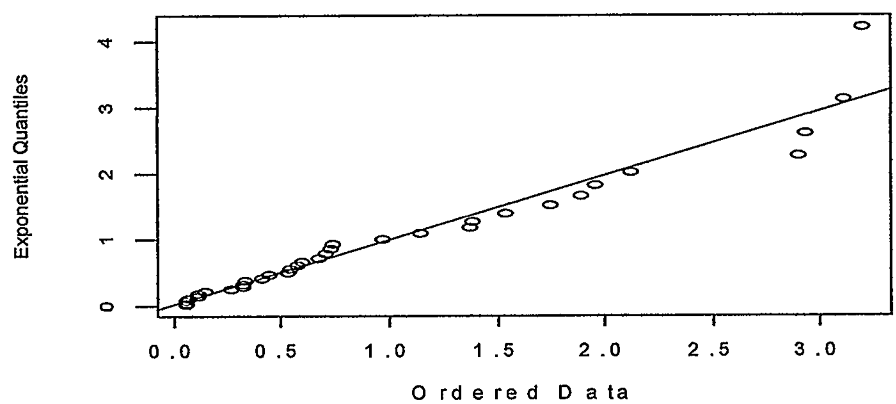


n = 80 (k = 81)

(b) Non-winter data



n = 10 (k = 366)



n = 110 (k = 33)

It is shown in Figure 3-4 that when the block size is n = 10, the residuals are not approximately exponential distributions, which is represented by points not on straight lines. When n ≥ 80, the residuals are approximately exponentially distributed with most points on straight lines, which means that the GEV model fits the data well.

3.3.3 Maximum Likelihood Estimations of GPD

A similar basic problem for GPD estimation is to choose a suitable value of threshold, u, a threshold shows the beginning of a tail distribution converging to GPD. Based on the Pickands-Balkema-de Haan Theorem, a high enough threshold u should be chosen in order that GPD fits the conditional tail distribution. The dilemma is that although a high threshold u ensures a tail distribution as GPD, it leads to less data of excceedances and thus high standard errors.

A tool used in the threshold determination is the Hill-plot. Hill, B. M proposed an estimator of ξ for $\xi > 0$ (Gencay and Selcuk, 2004). By ordering the sample data X_1, X_n with respect to their values as $X_{1,n} \geq X_{2,n} \geq \ldots \geq X_{n,n}$, the Hill estimator of the shape parameter ξ is

$$\hat{\xi} = \frac{1}{k} \sum_{l=j}^{k} \ln X_{j,n} - \ln X_{k,n},$$

where $k \rightarrow \infty$ is upper order statistics (the number of exceedances), and n is the sample size. Roughly speaking, the higher the value of ξ, the heavier the tail and thus the larger the quantile estimates.

In practice, there are two kinds of Hill-plots. A basic Hill-plot is plotted according to the Hill estimator function by fitting GPD with different thresholds to obtain the maximum likelihood estimates of ξ. The estimated shape parameter ξ is plotted on the y-axis. The number of data points exceeding the threshold is plotted on the lower x-axis and the thresholds is plotted on the upper x-axis. This Hill-plot shows a 95% asymptotic confidence interval of the estimation. In general, a threshold is chosen from a part of the Hill-plot where the values of the shape parameter are stable.

Another option of Hill-plot is to use the estimated quantiles as y-axis values, showing different thresholds related to different quantile estimates.

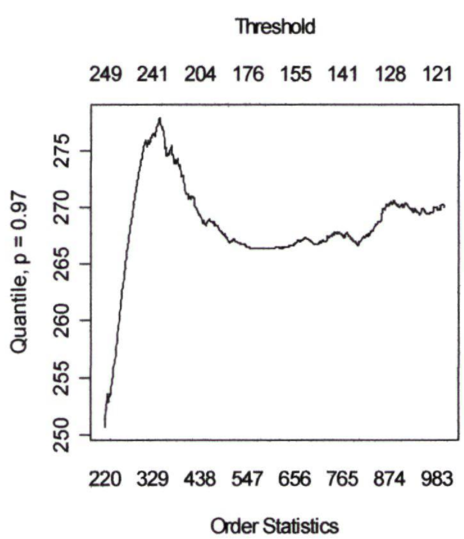
Figure 3-5 shows the two kinds of Hill-plots for the Alberta data. The Hill-plots of ξ estimates suggest that, for the winter data, a threshold is chosen between_values around u = 100 and u=250. For the non-winter data, the Hill-plot suggests a high threshold from u = 100 to u = 160.

The Hill-plots of quantile estimates show stable estimates of VaR (at p = 0.97) between u = 100 and u = 250 for the winter data, and between u = 100 and u = 160 for the non-winter data. Among the same threshold interval, the estimates of VaR (at p = 0.999) are not stable, showing more volatility in the extreme value estimations.

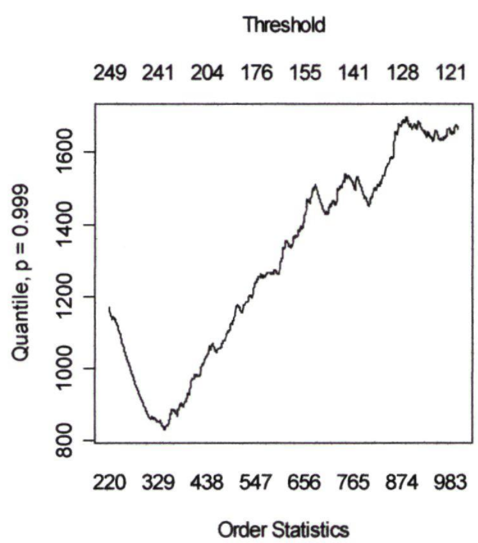
Figure 3-5 Hill plots

Part I Winter data

(a) Hill plot of xi estimates (CI = 0.95)



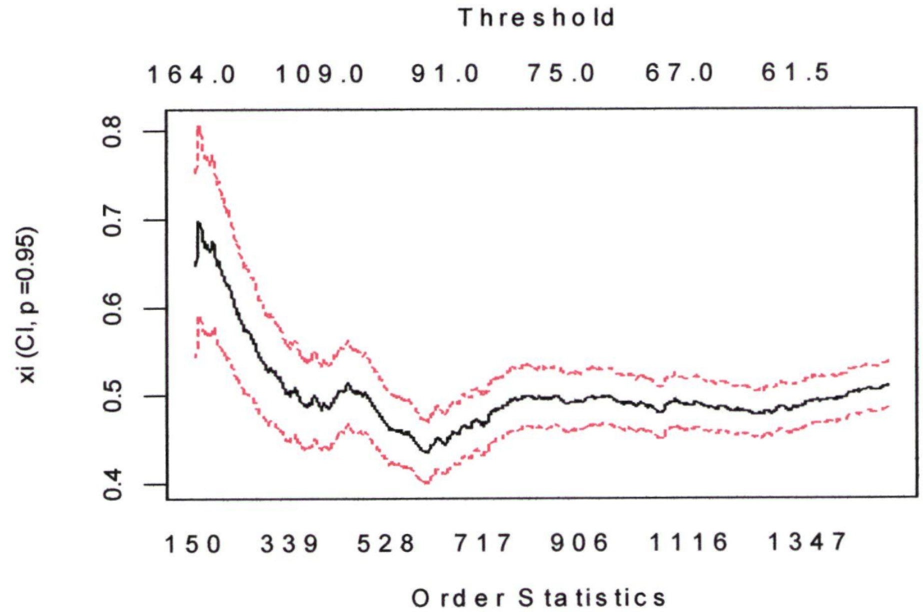
Quantile (p=0.97)



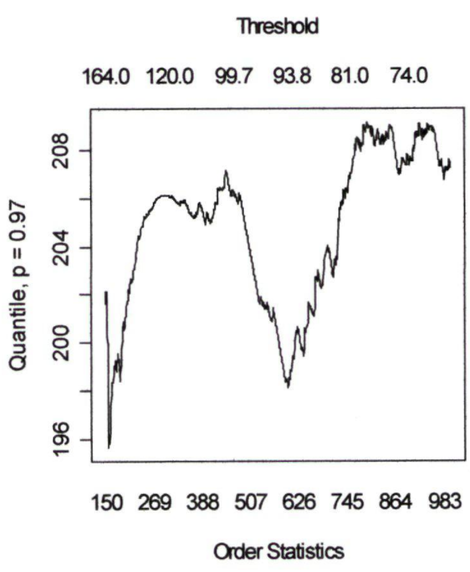
Quantile (p=0.999)

(b) Hill plot of quantile estimates

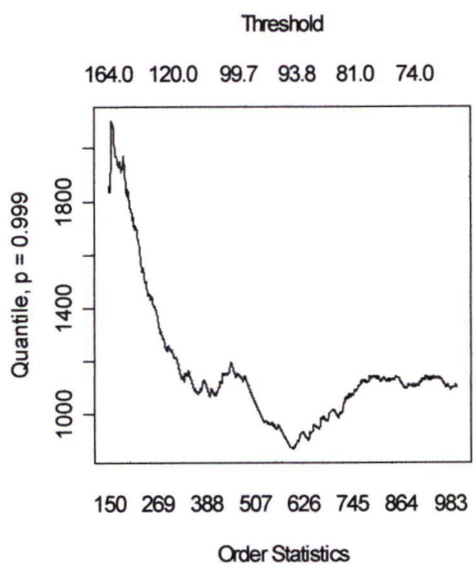
Part II Non-winter data



(a) Hill plot of xi estimates (CI = 0.95)



Quantile (p=0.97)



Quantile (p=0.999)

(b) Hill plot of quantile estimates

Table 3- 4 Parameter estimates for GPD models

Raw data		Winter (N = 7290)			Non-winter (N = 3660)		
Threshold	p	0.85	0.9	0.95	0.85	0.9	0.95
	u	111.33	145.73	230.16	96	114.02	141.38
Extremes	n	1094	729	365	547	366	183
GPD Estimates	ξ	0.187020	0.137743	0.517094	0.785705	0.932226	0.158997
	s.e (ξ)	0.0376274	0.0434214	0.120233	0.080409	0.1113176	0.09974
	β	87.510657	102.3862	66.38154	31.43644	34.5744	150.6797
	s.e.(β)	4.195649	5.818412	7.41507	2.661378	3.9673	18.6269

The empirical maximum likelihood estimates of the GPD parameters and the VaR and ES estimates for the Alberta data are shown in Table 3-4 and in Table 3-5. One can see from Table 3-4 that the standard deviation of the estimated parameter increases as the quantile of threshold increases. The confidence intervals for the VaR and ES estimates in Table 3-5 are asymmetric with higher upper bounds than lower bounds.

To see the fitness of GPD to the empirical data, one can use a plot of the excess distribution and QQ-plot of the residuals for a certain threshold value *u*. For the Alberta data, the threshold *u* is chosen at the percentile p = 0.95 for the GPD modeling. Figure 3-6 shows the plot of the excess distribution and the QQ-plot of the residuals.

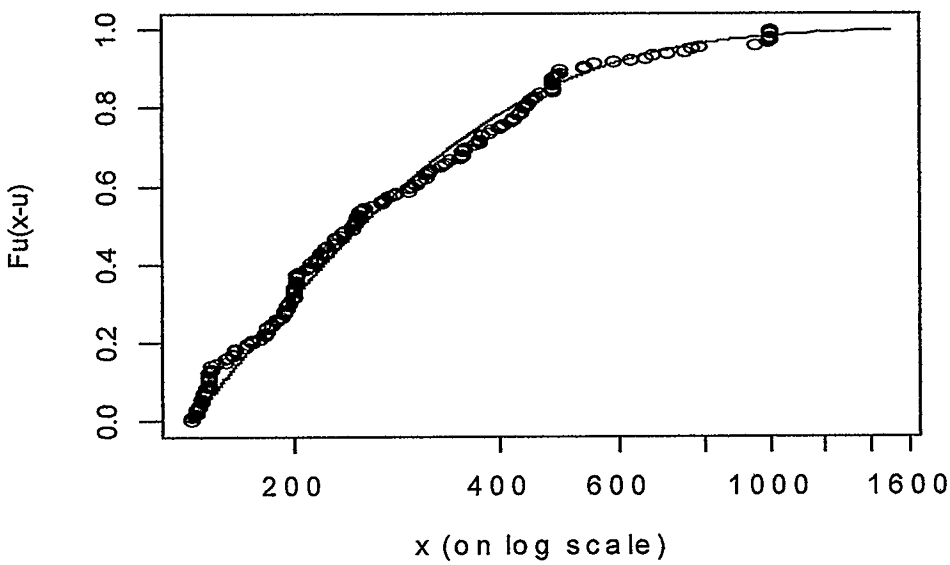
Plots in Figure 3-6 show good fitness of GPD for both winter and non-winter data. As a result, it is prudent to set the high threshold at p = 0.95 level. The quantile of the chosen threshold u at the high percentile q = 0.95 is 230.16 for the winter data, and 141.38 for the non-winter data. The VaR estimates above the chosen threshold at percentiles p = 0.97, 0.99, 0.999 are 269.0882, 397.0555, and 1072.9953 for the winter data, and 221.5587, 417.7387, and 958.8986 for the non-winter data. The ES estimates above the chosen threshold at percentiles p = 0.97, 0.99, 0.999 are 448.2352, 713.2293, and 2112.963 for the winter data, and 415.8845, 649.1536, and 1292.6231 for the non-winter data. The confidence intervals of these estimates are displayed in Table 3-5.

Table 3-5 GPD estimated VaR and ES

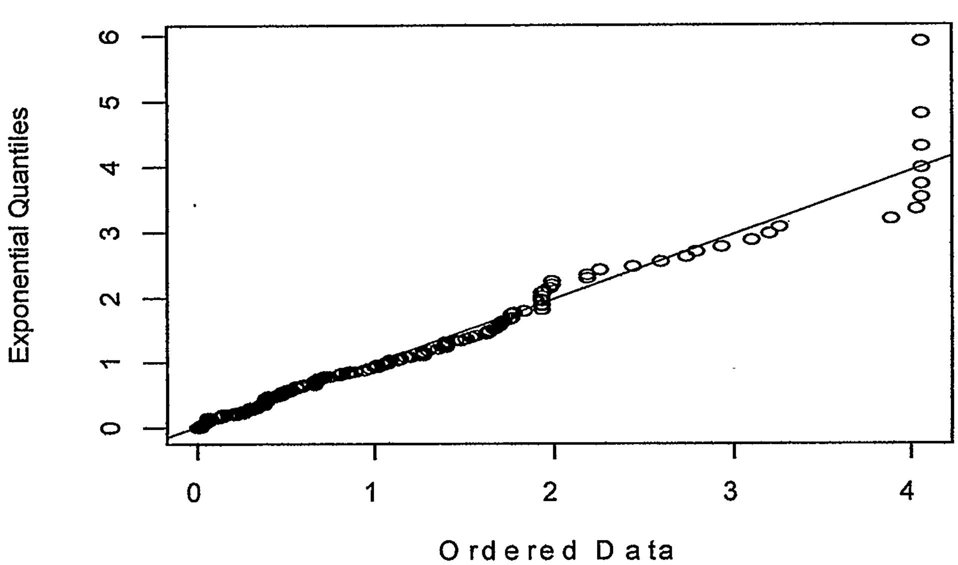
Raw data			Winter (N = 7290)			Non-winter (N = 3660)		
Threshold	p		0.85	0.9	0.95	0.85	0.9	0.95
	u		111.33	145.73	230.16	96	114.02	141.38
Extremes	n		1094	729	365	547	366	183
VaR (q) Estimates	q = 0.97	VaR (q)	275.7236	279.82	269.0882	197.2799	190.8727	221.5587
	CI	Upper CI	287.8089	291.9639	275.258	213.8676	206.0949	239.5646
		Lower CI	265.2524	270.5211	268.2153	183.7441	178.5305	206.2151
	q = 0.99	VaR (q)	419.9495	423.1652	397.0555	390.9467	394.2254	417.7387
	CI	Upper CI	448.3454	449.2374	425.2012	476.5997	492.8279	468.4331
		Lower CI	397.9516	401.0185	373.888	335.2481	334.1847	375.9047
	q = 0.999	VaR (q)	837.9023	804.1351	1072.995	2100.994	2791.405	958.8986
	CI	Upper CI	984.6927	937.067	1499.985	1498.5	NA	1339.638
		Lower CI	738.1378	715.2786	835.967	1343.593	NA	785.5563
ES (q) Estimates	q = 0.97	VaR (q)	421.1823	419.9812	448.2352	715.3161	1758.095	415.8845
	CI	Upper CI	455.939	450.8721	579.9811	1498.5	1498.5	484.935
		Lower CI	397.9516	396.2474	402.7008	456.909	532.315	373.9978
	q = 0.99	VaR (q)	598.5864	586.2254	713.2293	1619.055	4758.533	649.1536
	CI	Upper CI	676.3254	655.1367	1106.787	1498.5	1498.5	834.3869
		Lower CI	544.3207	538.8517	582.5228	875.9285	1110.773	561.6198
	q = 0.999	VaR (q)	1112.686	1028.054	2112.963	NA	NA	1292.623
	CI	Upper CI	1406.857	1287.898	1499.985	NA	NA	1498.5
		Lower CI	928.799	870.6765	1260.385	NA	NA	965.6855

Figure 3-6 Plots of GPD fitness

The Excess Distribution Plot (u = 230.16)

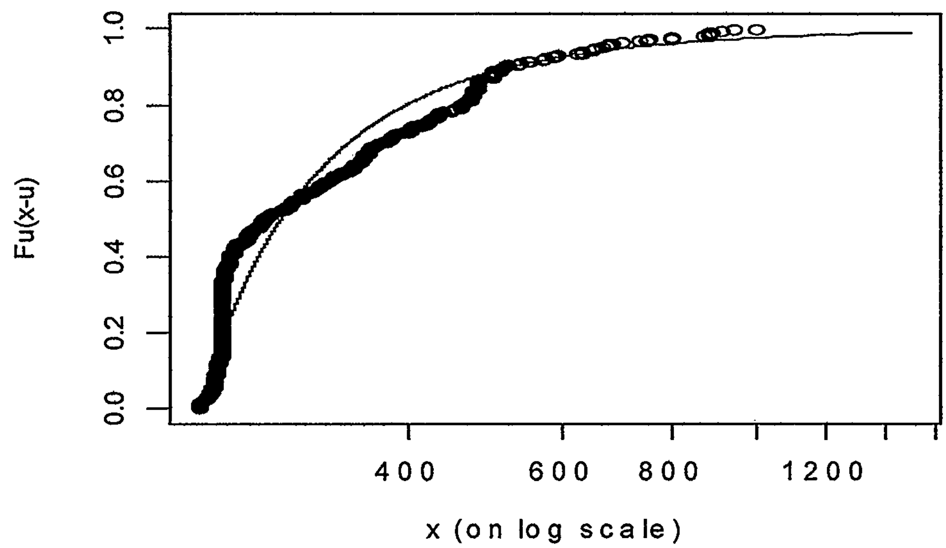


QQ-plot of Residuals (u = 230.16)

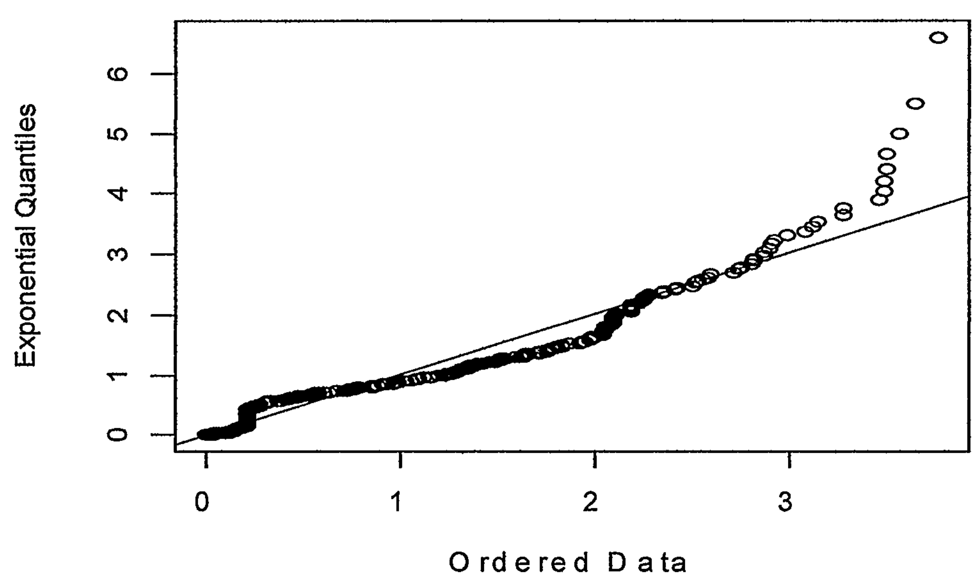


(a) Winter data

The Excess Distribution Plot (u = 141.38)



QQ-plot of Residuals (u = 141.38)

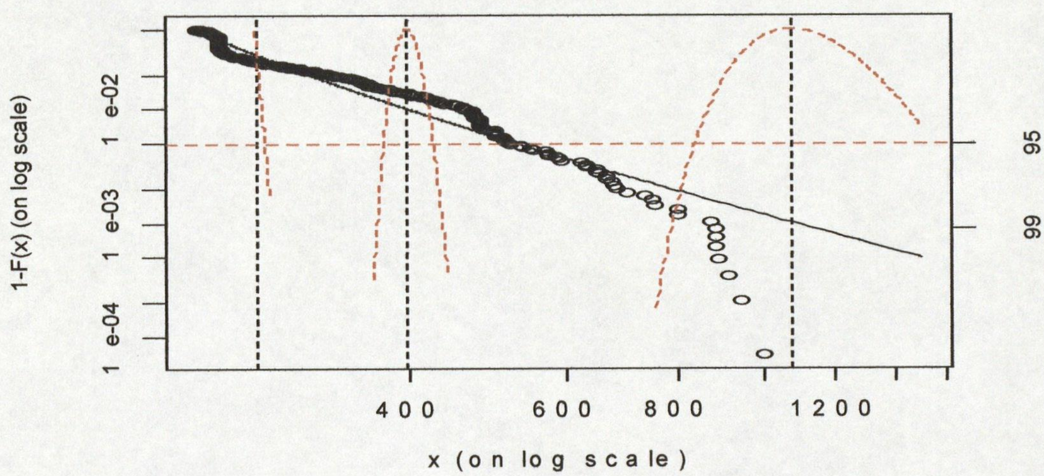


(b) Non-winter data

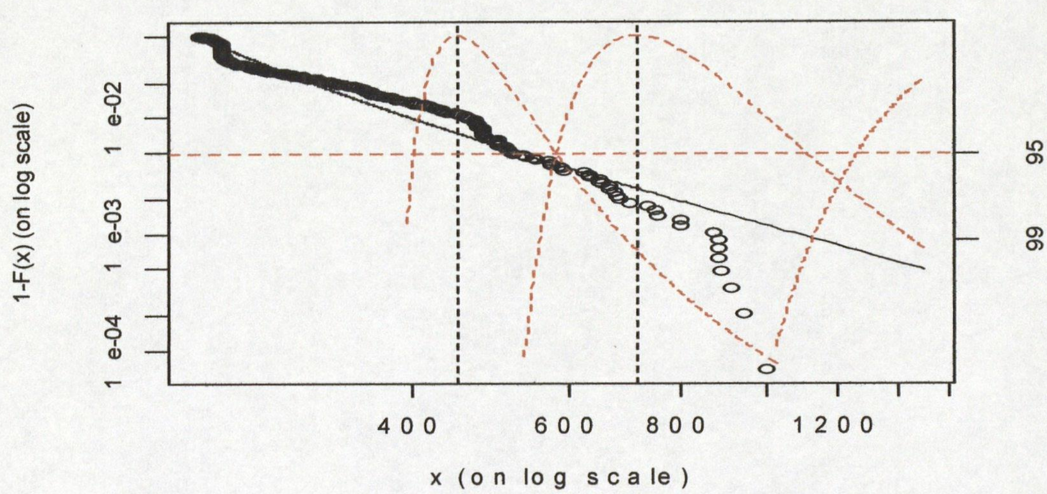
Figure 3-7 shows the VaR and ES estimates and their 95% confidence intervals.

Figure 3-7 GPD estimated VaR and ES

Part I Winter data with the threshold $u = 230.16$

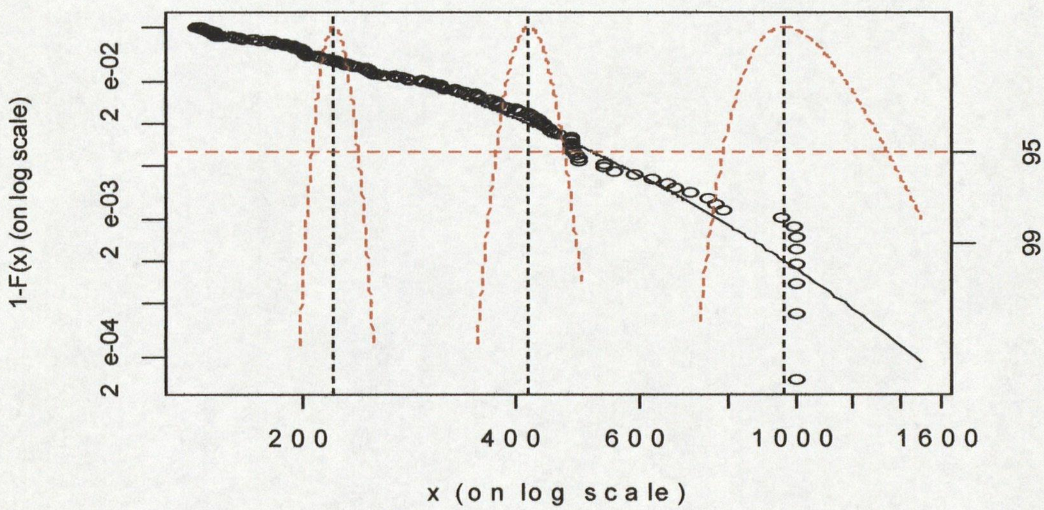


(a) VaR estimates at quantile 0.97, 0.99, and 0.999

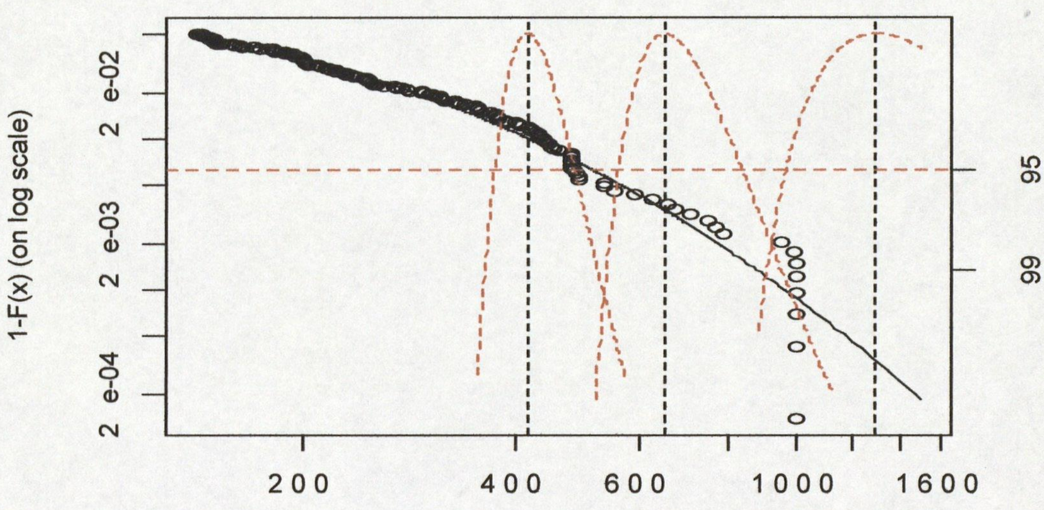


(b) ES estimates at quantile 0.97, 0.99, and 0.999

Part II Non-winter data with the threshold u = 141.37



(a) VaR estimates at quantile 0.97, 0.99, and 0.999



(b) ES estimates at quantile 0.97, 0.99, and 0.999

4. Performance of VaR Models

4.1 VaR Estimates from Different Models

In the above chapter, the estimations of VaR from GEV and GPD models are based on the extreme value theory (EVT). Chapter 2 introduces other traditional methods for VaR estimation: the variance-covariance method with normal distribution model and the historical simulation method. To see which method is the most accurate, it is necessary to compare VaR estimates from EVT models and from the traditional methods.

Table 4-1 shows the estimates of VaR from four different models at high percentiles q = 0.97, 0.99, 0.999. The chosen block size of GEV estimates of parameters and VaR is n = 90 for the winter data, and n = 110 for the non-winter data. The chosen threshold for GPD estimates of parameters and VaR is q = 0.95 for both the winter and non-winter data. In the normal distribution model, there is no parameter estimation, and VaR are directly estimated from the symmetric distribution with the sample mean and standard deviations. In the historical simulation method, with no distribution assumptions and thus no parameter estimation, VaR are simulated from the realized historical data.

It is obvious that different methods give different estimation results. Table 4-1 shows that the VaR estimates from GEV are much larger than VaR estimates from other models. Recall from chapter 3, when choosing a large block size (n = 80 for the winter data, n = 110 for the non-winter data) to fit GEV to the standardized maximum distribution, there is a big loss of observations in the data set. As a result, although the residual plots show good fitness of GEV for the maxima, with few observations of maxima (k = 81 for the winter data, k = 33 for the non-winter data) the VaR estimates from GEV have large confidence intervals. Thus, this study will avoid using the estimation results from GEV for comparison. The EVT estimation method is represented by GPD in the comparison with other traditional methods.

Table 4-1 VaR estimates from different models

Raw data	Model	Blo-size	Maxi	Thres	Shape	VaR (q) estimates		
		n	k	q(u)	ξ	q = 0.97	q = 0.99	q = 0.999
	GEV	90	81		0.12139	954.564	1278.23	2108.94
Winter	GPD			0.95	0.51709	269.0882	397.0555	1072.995
N=7290	Nor					230.2404	256.684	326.4507
	Hist					250	457.7096	847.8198
	GEV	110	33		0.2057	1161.75	1647.52	3094.14
Non-winter	GPD			0.95	0.159	221.5587	417.7387	958.8986
N=3660	Nor					214.4083	249.2583	308.8837
	Hist					211.925	434.9273	998.9341

Comparing VaR estimates from the GPD model and from the normal distribution model in Table 4-1, it is clear that at each high percentile q = 0.97, 0.99, 0.999, VaR estimates from the GPD model are larger than that from the normal distribution model. Comparing VaR estimates from the GPD model and those from the historical simulation, at q = 0.97, 0.999 the GPD estimates of VaR are larger than the historical estimates. Whereas at q = 0.99, the historical estimates of VaR are larger than the GPD estimates. If the GPD estimates are closest to the true values, the normal distribution model will always underestimate the risks of high quantiles, and the historical simulation method will some times underestimate the risks of high quantiles.

In fact, there is no proof that GPD estimates are closest to the true values yet. To evaluate which model estimates the most accurately, a back-testing method is introduced to compare the performance of various models.

4.2 Back-testing Process and the Comparison of Estimation Models

4.2.1 Back-testing Process

With the development of modeling methods for market risk analysis, many banks have adopted a process to test the quality and accuracy of their risk measurement systems. The process is the back-testing process, summed up as a framework by Basle Committee on Banking Supervision in 1996 (Basle Committee, 1996a). In a bank system, the basic idea of the back-testing process is to compare actual trading results with the model-generated risk measures. If the result is close, there is no systematic problem in the risk measurement modeling. If the result is very different, there may be some problems in the modeling system.

A practical application of the back-testing process is to use a "sliding time window". Gencay et al. (2004) use this process to test the VaR estimates with the data from the investment returns in emerging markets. McNeil and Frey (2000) use this process to test the VaR estimates with the BMW data. The follow expression about the back-testing process is based on Gencay et al. (2004) and McNeil & Frey (2000).

Suppose there is a historical series with N observations. To back-test the estimates from risk measurement models, one chooses a subset data size m, where $m \ll N$. The sub-data set is called a time window of m observations. If one places the first window between the 1^{st} and the m^{th} sub-data set, one can estimate the $(m + 1)^{th}$ VaR_q at a high quantile q by this m sub-data set modeling, and then compare the estimates with the real $(m + 1)^{th}$ observation value. When moving the window one period ahead from the 2^{nd} to the $(m + 2)^{th}$ sub-data set, it is called sliding a window. The estimate of the $(m + 2)^{th}$ VaR_q is from another m sub-data set modeling and thus has a different value. This time the comparison is between the estimated $(m + 2)^{th}$ VaR_q and the realized $(m + 2)^{th}$ observation value. Repeating the above steps, one can slide a time window of m observations in a set of historical data with N observations, getting $(N - m)$ estimates of VaR_q from the back-testing process.

The performance of a risk measurement model is summarized by a violation value (violation number or violation ratio). A violation happens when the realized observation value is greater than the related estimated VaR_q. By summing all the violations among $(N - m)$ times of estimations, one gets the violation number of a back-testing process. A violation ratio is the ratio of a violation number divided by the total number of estimations $(N - m)$ during a back-testing process.

To evaluate the performance of a model, a critical value becomes the "expected violation value" (expected violation number or expected violation ratio). When estimating VaR at a certain high percentile *q*, one expects that the realized observation values will have *(1 – q) N* times or *(1 – q)* percent of the times be higher than the estimated VaR_q, if the model is correct. In other words, the expected violation number is *(1 – q) N* and the expected violation ratio is *(1 – q)*. For an example of the violation ratio, if one chooses percentile q = 0.97, 0.99, 0.999 for VaR_q estimations, the expected violation ratios are (1 – q) = 0.03, 0.01, 0.001.

The performance of a model is evaluated by comparing the realized violation value with the expected violation value at a given high percentile. If the realized violation number (or ratio) is larger than the expected one, it means that the model consistently underestimates VaR in the upper tail of the underlying distribution. If the realized violation number (or ratio) is less than the expected one, it means that the model consistently overestimates VaR in the upper tail of the underlying distribution. As a result, if all models underestimate VaR, the model with the least underestimation is most accurate. If all models overestimate VaR, the model with the least overestimated is most accurate. If some models overestimate VaR and other models underestimate VaR, one compares the absolute difference level of the overestimates and underestimates. The model with the least absolute difference is the most accurate.

Whether a small or a large violation value is desirable depends on the perspectives of different organizations. In a banking system, regulatory organizations may prefer a model with a smaller violation value, since this model consistently overestimates risks and thus requires more capital allocation than necessary. As a result, an adequate amount of capital will be allocated in the banking system in case of excessive losses. On the other side, commercial financial institutions may prefer a model with a larger violation value, since this model consistently underestimates risks and requires less capital allocation than necessary. The result is that less capital is set aside preparing for extreme situations, and more capital is available for commercial investments.

In a deregulated electricity wholesale market, high prices are good for generators and bad for retailers and self-retailers. However, it is not known for either side of the business whether a larger or a smaller violation value is preferred. As a result, one may consider that they both prefer the exact accurate of the violation values. This paper compares absolute differences between violation values and their expected violation values, and concludes that the model with smallest absolute difference is most accurate and desirable.

4.2.2 Comparison of the Back-testing Results

In the back-testing processes, this study chooses two different sizes of the sliding window for both the winter and non-winter data set. One window size is 1000, which means using a data set with 100-day peak hour prices to predict the 1001st peak hour price. The other window size is 1500, which means using a data set with 150-day peak hour prices to predict the 1501st peak hour price.

These window sizes are chosen not to be too small, so that a high percentile estimate of VaR can be tested. For example, when estimating VaR at percentiles q =

0.97, 0.99, 0.999 with window size 1000 and 1500, the expected violation numbers are 30, 10, and 1 for window size 1000, and 45, 15, and 1 (or 2) for window size 1500. To test the estimation results at the high percentile q = 0.999, a window size of at least 1000 is necessary.

On the other hand, the window size is chosen not to be too large, so that enough times of estimation are included in a back-testing process. For example, for the non-winter data set with 3660 observations, when choosing window size 1000 and 1500, one will back test the estimation 2660 and 2160 times individually during the back-testing process.

Violation numbers and violation ratios from the back-testing process for the GPD model, the normal distribution model, and the historical simulation method are shown in Table 4-2 and Table 4-3. The performance of each model is shown by the differences between the real violation values and the expected violation values. The differences are values in the "error" columns in Table 4-2 and Table 4-3.

From Table 4-2, at the percentile q = 0.97, all the differences between real and expected violation numbers from three different models are positive, which means that all models consistently underestimate risks. At the percentile q = 0.99, two of the differences between the real and expected violation numbers from the GPD model and the historical simulation are negative for the non-winter data with the window size 1500. This means that these two models consistently overestimate risks. At the percentile q = 0.999, the back-testing process for the non-winter data with the window size 1500 also shows overestimations of VaR from the GPD and historical models. Despite the above exceptions, all the other testing processes at quantiles 0.99 and 0.999 show positive differences between the real and expected violations.

In Table 4-2, the expected violation number for the winter data at the window size 1000 is 189, while the real violation number from the GPD estimation is 191, from the normal distribution estimation is 221, and from the historical simulation estimation is 198. The least difference between the real violation number and the expected violation number is 2 from the GPD estimation. Thus, the conclusion from this back-testing process is that GPD is the most accurate model for the VaR estimation at the high percentile $p = 0.97$. The same conclusion can be derived from the winter data with the window size 1500 and from the non-winter data with the window size 1000. The exception at the $p = 0.97$ estimation is from the non-winter data with the window size 1500. Here the least violation number 8 is from the normal distribution model. In summary, for VaR estimations at the high percentile $p = 0.97$, four back-testing processes show that the two most accurate results come from the GPD model, one from the normal distribution model, and one from both the GPD and normal distribution models.

Table 4-2 Violation numbers from different models

Season		Oct-May (winter)				June-Sep (non-winter)			
			Error		Error		Error		Error
Data	N	7290		7290		3660		3660	
Window	m	1000		1500		1000		1500	
Estimations	(N – m)	6290		5790		2660		2160	
	Expected	189	0	174	0	80	0	65	0
p=0.97	GPD	191	2	180	6	99	19	76	11
	Normal	221	32	210	36	99	19	73	8
	Historical	198	9	190	16	102	22	82	17
p=0.99	Expected	63	0	58	0	27	0	22	0
	GPD	83	20	66	8	50	23	12	-10
	Normal	179	116	165	107	87	60	55	33
	Historical	76	13	65	7	50	23	11	-11
p=0.999	Expected	6.3	0	5.8	0	2.7	0	2.2	0
	GPD	8	1.7	9	3.2	4	1.3	1	-1.2
	Normal	129	122.7	123	117.2	64	61.3	39	36.8
	Historical	15	8.7	11	5.2	7	4.3	0	-2.2

Similarly, for VaR estimates at the percentile $q = 0.99$, four back-testing processes show that one of the most accurate results from the GPD model, two from the historical simulation, and one from both the GPD model and the historical simulation models. For VaR estimations at the quantile q = 0.999, four back-testing processes show that all the most accurate results are from the GPD model.

An overall conclusion from Table 4-2 is that, compared with the other two models, the GPD model is the most accurate one for extreme value estimations for q = 0.999. The performance of the normal distribution is poor at any high percentile in the tail of underlying distributions (in our analysis a tail begins from q = 0.95). The performance of the historical simulation is better than that of the normal distribution at any high quantiles, and sometimes even better than the GPD model at a modest high quantile (q = 0.99). This is reasonable since a normal distribution is not suitable for asymmetric and fat tailed distributions, and a historical simulation captures modest high values but not extreme values. The GPD model fits any distribution, including fat-tailed and asymmetric. The same conclusion is derived when comparing the difference of the real and expected probabilities in Table 4-3.

Table 4-3 Violation ratios

Season		Oct to May (winter)				June to Sep (non-winter)			
			Error		Error		Error		Error
Data	N	7290		7290		3660		3660	
Window	m	1000		1500		1000		1500	
Estimates	n	6290		5790		2660		2160	
p=0.97	Exp	0.0300	0	0.0300	0	0.0300	0	0.0300	0
1-p=0.03	GPD	0.0304	0.0004	0.0311	0.0011	0.0372	0.0072	0.0352	0.0052
	Nor	0.0351	0.0051	0.0363	0.0063	0.0372	0.0072	0.0338	0.0038
	His	0.0315	0.0015	0.0328	0.0028	0.0383	0.0083	0.0380	0.0080
p=0.99	Exp	0.0100	0.0000	0.0100	0.0000	0.0100	0.0000	0.0100	0.0000
1-p=0.01	GPD	0.0132	0.0032	0.0114	0.0014	0.0188	0.0088	0.0056	-0.0044
	Nor	0.0285	0.0185	0.0285	0.0185	0.0327	0.0227	0.0255	0.0155
	His	0.0121	0.0021	0.0112	0.0012	0.0188	0.0088	0.0051	-0.0049
p=0.999	Exp	0.0010	0.0000	0.0010	0.0000	0.0010	0.0000	0.0010	0.0000
1-p=0.001	GPD	0.0013	0.0003	0.0016	0.0006	0.0015	0.0005	0.0005	-0.0005
	Nor	0.0205	0.0195	0.0212	0.0202	0.0241	0.0231	0.0181	0.0171
	His	0.0024	0.0014	0.0019	0.0009	0.0026	0.0016	0.0000	-0.0010

5. Risk Management Application in the Power Market

After discussing the unique properties of electricity and the structures of the restructured power market, this chapter gives an application of using the extreme value theory in evaluating hedging options on the spot power market.

5.1. Electricity Properties and Power Market Structures

Electricity, the commodity traded in power markets, appears to be a perfect commodity, since all electrons are naturally identical. In fact, the unique non-storable characteristic of electricity makes the power delivered at any particular time to any particular location a different commodity. One cannot buy power early, store it and then sell it later for higher prices. The non-storable feature of electricity requires the real-time balancing of supply and demand system-wide. In other words, the instantaneous supply and demand must be in balance. Otherwise, the reliability of the whole power system may be impacted. This peculiar feature of the electricity market introduces the need for an additional set of services, beyond production and distribution, the balancing and reserve resources. Therefore, the supply of electricity involves three types of activities: generation, transmission, and ancillary services or balancing (Eydeland and Wolyniec, 2003).

In deregulated power industries, the primary cash market takes on two contracting structures: pools and bilateral markets. The main characteristic of the pool market is the formal establishment of the market (system)-clearing price at which all cash (energy) transactions clear. In the bilateral markets, all transactions are made by two parties and are independent of any other transactions in the market.

In Alberta, the format of the primary cash market is the Alberta Power Pool. It is an hourly wholesale spot market, operated by the Alberta Electric System Operator (AESO). AESO is independent of any industry affiliations and owns no transmission or market assets. AESO provides open transmission accesses to the Alberta Interconnected Electric system (AIES) for generators, distribution companies and large industrial consumers. AIES also provides contracts with transmission facility owners to acquire transmission services and provide customer accesses. Consistent with its responsibility to ensure system reliability, AESO procures ancillary services, including operating reserves, to address contingencies and moment-to-moment changes in load (AESO, "Fast Facts").

In addition to spot power markets, forward power markets are markets where the parties contract for the delivery of power in the future. There are three basic forms of forward power markets: bi-latered or broker-based (over-the-counter), market maker-based, and exchange-based. In an over-the-counter forward market, the trading involves either direct contract between two parties or a contact mediated by a broker. In a market maker-based forward market, the trading is centered on a market maker who posts two-sided (buying and selling) quotes, stands behind every transaction, and can carry inventory. In an exchange-based forward market, a central exchange matches up buyers and sellers, and guarantees the performance of the transaction without taking an outright position and carrying inventory (Eydeland and Wolyniec, 2003).

In Alberta, an exchange-based forward market is operated by the Alberta Watt Exchange Limited (Watt-Ex). Watt-Ex is an online trading system for the electrical energy in Alberta. It allows market participants to buy or sell contracts for firm forward commitments in the real time or spot electricity market, as well as trading of ancillary services. Through this forward market, buyers and sellers can lock in a firm price for electricity, or arrange direct sales or forward market contracts for their electricity needs (http://www.watt-ex.com).

5.2 Application of the GPD Estimated VaR: Hedging Power Price Risks with Call Options

It is well known that electricity prices are highly volatile, due to the variation in demand, together with the constraint in generation capacity. Many derivative instruments have been developed for the risk management of energy transactions. Derivatives are financial securities whose value is derived from another underlying financial security. Derivatives can be used in hedging, protecting against financial risks, or can be used to speculate on the movement of security or commodity prices, or the levels of financial indices. The valuation of derivatives is based on the statistical mathematics of uncertainty. In power markets, the non-storability of power is also a reason why well-known financial theories may not be applicable to electricity derivatives.

5.2.1 Plain-vanilla Options

There are several standard derivative products, often called "plain vanilla", used as risk management tools in energy markets. Futures contracts, forward contracts, swaps, and options are the most frequently used ones.

Futures are highly standardized exchange-traded contracts for the purchase or sale of an underlying commodity or financial product at a specified price over a certain future period. Forward contracts are agreements to buy or sell a commodity at a future time. When the forward contract is made, the price to be paid at delivery is specified. Swaps in energy markets are similar to swaps in financial markets and are natural generalizations of forward contracts (Eydeland and Wolyniec, 2003).

After futures, forwards, and swaps, the other most frequently used risk management tools are plain-vanilla options. The plain-vanilla options are standard for the power and natural gas markets, including calls and puts. In energy markets, a call option is the right, but not the obligation, to buy energy at a predetermined strike price, and a put option is the right, but not the obligation, to sell energy at a predetermined strike price. Typically, energy option specifications include location, exercise time, delivery conditions (for example the type of delivered power, such as on-peak, off-peak, round-the clock), strike, and volume (Eydeland and Wolyniec, 2003).

Based on the exercise time, there are two kinds of options: European-style options, which exercised only once at the specified exercise date, and American options, which can be exercised only once at any time before the exercise date. In current energy markets, there are two groups of popular options: physical options and options on the spot market.

a. Physical Options

To see the payoff at an option exercise, one needs to separate the financially settled options from the physically settled options. The usual exercise periods of the options in current energy markets are calendar year, quarter, and month.

For financially settled options, there must be a widely accepted financial index (such as Inside FERC' Gas Market Report in the gas market). The options are exercised against the index. If K is the strike price, and S is the price of the index against which the option is settled, the call's payoff of a European option is $\Pi_{call} = \max\{S - K, 0\}$, and the put's payoff of a European option is $\Pi_{put} = \max\{S - K, 0\}$.

Physically settled options are power market instruments since power markets have not developed a financial index for the financial settlement. The holder of a physically settled call (or put) option has the right, but not the obligation to buy (or sell) the commodity for the contracted period, paying the strike price K. To extract the value from a physically settled European option, the second step for the option holder is to sell (or buy) the commodity at the spot market. The spot market price S used in exercising the option is the average spot price inside the contractual period. Again, the payoff of the option is $\Pi_{call} = \max\{S - K, 0\}$ for call options, and $\Pi_{put} = \max\{S - K, 0\}$ for put options.

b. Options on the Spot Commodity

Usually, options on spot commodity gas or power are excised daily during a certain period, such as month, quarter, and season. Two groups of options on the spot commodity exist in the energy market: fixed-strike options and floating-strike options. The owner of fixed-strike daily option can make daily decisions during the exercise period about buying (call option) or selling (put option) spot commodity at a fixed strike price. The owner of a floating-strike daily option can make daily decisions during the exercise period about buying or selling spot commodity at a strike price determined at the beginning of the exercise period as a settled value of the period's index.

Typically, in the power market, hourly options on the spot commodity are available and used to manage price risks on an hourly basis. They are options financially settled against the real-time hourly prices such as power pool prices in Alberta.

5.2.2 The Application Example: Hedging with Options as Insurance against Extraordinarily High Power Prices

Because of the unpredictable demand from the consumer side, there are situations when a retail company must buy power at on-peak hours from the wholesale spot market to fulfill their obligations of supplying enough power to their customers. In this case, the retailers want to protect themselves from extraordinarily high prices in the spot wholesale market, since the high price spikes are typically low-probability events with possibly destructive effects on the company. In the above situation, an hourly option on the wholesale spot power market can be used as insurance for the retail companies. The details of how the option works are explained in the follow example.

Suppose during the winter season (from October to May) in Alberta, a retail company has to offer some unexpected extra power to its customers during some period of on-peak hours. In addition to the contracts in the forward market with power generators, it has to buy the extra power from the spot wholesale market. Since the high amount of demand in power always relates to high system margin prices, the retailer wants to find a way of hedging against the possible extremely high prices.

Suppose the only available call option for this insurance purpose is an hourly option as follows: 100 MW Calgary on-peak Winter 2004-2005 $200 Call with premium $6/MW per day.

This is an option financially settled against the real-time hourly prices of the Alberta Power Pool. The owner of this option has the right, but not the obligation, to call on the seller of the option to deliver 100 MWh of power at $200 /MWh in the city of Calgary. The contracted period is the on-peak hours (from 11 AM to 20 PM in this case) during the winter of 2004-2005 (from October 2004 to May 2005 in this case)

The call option buyer has to pay $6 /MW per day during the eight month from October 2004 to May 2005. Since the market participants are more willing to buy calls than selling them, especially when prices are high due to the constraint of supply, the high premium is unavoidable. It is clear that the total premium during the exercise period is:

$$C = \$6 \text{ (/MW per day)} * 243 \text{ (day)} * 100 \text{ (MW)} = \$145{,}800$$

Now the retailer has to decide whether it should buy the call option or not. Suppose the company only considers the 0.999 quantile level as extremely high prices, which means that there are 2 hours (= 0.001 * 2430 h) during the eight-month (2430 hours) period when the extremely high prices occur. Based on different estimation models, different retailer companies will have different VaR estimates at the same quantile level. Thus, they end up with different decisions about whether to buy the option or not.

The VaR estimates predicted by the normal distribution model, the historical simulation model, and the GPD model at the 0.999 quantile level are \$326 /MWh, \$848 /MWh, and \$1073 /MWh separately. Noting that in the Alberta Power Pool, the highest price is capped at \$1000 /MWh, this example will use \$1000 /MWh instead of \$1073 /MWh as the VaR estimate at the quantile 0.999 from the GPD model. The different estimates of VaR are believed by different retailers as the true extremely high prices.

Let S_1, S_2 and S_3 be the total savings from exercising the option with VaR estimated by the GPD model, the normal distribution model, and the historical simulation model. A retailer will buy the option if the total saving is positive from exercising the option.

If the retailer buys the option, it must pay the total amount of the premium of \$145,800. Thus, by exercising the option for two hours, the total savings are calculated as follows:

$$S_1 = (\$1000 - \$200) \text{ /MWh} * 2 \text{ h} * 100 \text{ MW} - \$145{,}800 = \$14{,}200$$

$$S_2 = (\$326 - \$200) \text{ /MWh} * 2 \text{ h} * 100 \text{ MW} - \$145{,}800 = - \$120{,}600$$

$$S_3 = (\$848 - \$200)\ /\text{MWh} * 2\ \text{h} * 100\ \text{MW} - \$145{,}800 = -\ \$16{,}200$$

Based on the above calculation, a retail company would not buy the option if it estimated VaR at the 0.999 quantile by the normal distribution model or the historical simulation model, since the total savings S_2 and S_3 are negative. A retail company would buy the option if it estimated VaR at the 0.999 quantile by the GPD model, because the total saving S_1 is positive.

Recall in chapter 4, the performance of the three models in estimating VaR has been compared by using the back-testing method. For the winter data, all the models underestimate the real VaR. Among them, GPD is the least underestimated and the most accurate model. Keeping this in mind, one should believe that the realized extreme prices or the real VaR are most close to the GPD estimates.

As a result, when the extremely high prices in the market realized at $1000 /MWh for only two hours, the retailer without hedging by the call option would in fact lose money at an amount of L_1.

$$L_1 = (\$200 - \$1000)\ /\text{MWh} * 2\ \text{h} * 100\ \text{MW} = -\ \$160{,}000$$

Here the retail company will lose money by paying spot prices of $1000 / MWh instead of the strike price of $200 /MWh, although it does not pay $145,800 of the premium.

From this example, it is clear that when evaluating savings from hedging by call options on the spot power market, different estimates of VaR lead to different calculations of the total savings. If a retail company correctly uses the GPD model to estimate VaR at the 0.999 quantile, it will choose to buy the call option and consequently save $14,200. If the retail company uses traditional estimation models such as the normal distribution model or the historical simulation model, it will miscalculate the estimated VaR and choose not to buy the call option. As a result, when the market realizes the extremely high prices as predicted by the GPD model, the retail company will lose $160,000. The amount of loss is much larger than that of saving. This is why we consider the call option as a tool of hedging as insurance.

6. Conclusion

This paper introduces extreme value theory as an analytical tool in estimating electricity market risks. Market risks in a deregulated wholesale spot power market are extremely high (or some times low) market prices measured by Value at Risk (VaR) and the expected shortfall (ES). Extreme value theory is a study of the tails of various distributions. There are two groups of models in the extreme value theory: the block maximum models that study the fluctuations of sample block maxima, and the peaks-over-threshold models that study the tails over a high threshold of the underlying distribution.

Using data on Alberta Power Pool prices, this paper estimates VaR by the extreme value theory method and the other two traditional methods, the variance-covariance model with a normal distribution assumption and the historical simulation model. A back-testing process is used to compare the VaR estimations from these three methods. The comparison shows that the extreme value theory method is the most accurate among all three methods. This is not a surprise since the exploratory analysis of the Alberta Power Pool data shows a fat-tailed and asymmetric distribution. Based on these factors, a normal distribution model is biased because it has an exponential-tail and a symmetric distribution, while a historical simulation model is limited to the historical observations and gives no possibility of realizing values beyond the historical extreme values. Only the extreme value theory models are able to encompass the fat-tailed and asymmetric distributions, and also provide parameter estimates of the underlying distributions using the limited realizations of extreme values.

The application of the extreme value theory to risk management analysis in power markets is very broad. One of the applications is to evaluate hedging call options in a wholesale spot power market. When calculating savings from exercising the hedging call options, different models lead to different estimates of the total saving. Only with the estimates from the extreme value theory model would a retailer make the correct decision to buy the hedging options, and thus avoid big losses under extreme situations.

The contribution of this paper is to apply the statistical tool of extreme value theory to the risk management analysis in the restructured power market. With the ongoing movement of restructuring electricity industries around the world, more and more volatility in market prices will require further studies to find new risk management tools.

The conclusion of this paper is that the extreme value theory is a desirable statistical tool in analyzing the extremely high price risks in the Alberta electricity wholesale market. I am not sure whether or not the same tool is suitable to other energy market analyses at this point. To have more information about this, more studies need to be done.

REFERENCES

"Reform of the Electricity Industry in Australia", http://www.efa.com.au/reform.html

"Restructuring the U.S. Electric Power Industry", http://www.nei.org/doc

AESO, "Fast Facts", www.aeso.ca

Artzner, P., Delbaen, F., Eber, J. & Heath, D. (1997), "Thinking coherently", RISK, 10 (11), pp. 68-71

Balkema, A. & de Haan, L. (1974), "Residual life time at great age", Annals of Probability 2, pp. 792-804

Basle Committee on Banking Supervision (1996a), "Supervisory framework for the use of 'back-testing' in conjunction with the internal models approach to market risk capital requirements", BCBS Publications, http://www.bis.org/publ/bcbs22.pdf

Basle Committee on Banking Supervision (1996b), "Overview of the Amendment to the Capital Accord to Incorporate Market Risk", BCBS Publications, http://www.bis.org/publ/bcbs23.pdf

Blanco, C. (2001). "Calculating and using Value at Risk" Global energy Business, May/June 2001, pp. 13-15

Borenstein, S. (2002), "The Trouble with Electricity Markets: Understanding California's Restructuring Disaster", Journal of Economic Perspectives, 16 (1), pp.191-211

Carrillo, S., Croissant, O., Comezana, G., Escobar, M., Fernandez, P., Hernandez, N., Sanchez-Calle, A., Seco, L., and Suarez, A. (2002) "Extreme value theory in non-gaussian Mark-to Future calculations" AQR: March Issue

Doucet, J.A. (2004), "Electricity Restructuring: Will the Promise be fulfilled?" CERI Electricity Conference, Dec 07, 2004

Dowd, K. (1998). "Beyond Value-at-risk: The new science of risk management", John Wiley & Sons Ltd.

Duffie, D. & Gray, S., 1995, "Volatility in Energy Prices", Managing Energy Price Risks, pp.39

Embrechts, P. (2000). "Extreme value theory: Potentials and limitations as an integrated risk management tool." Derivatives Use, Trading & Regulation 6, 449-456

Embrechts, P., Kluppelberg, C., & Mikosch, C. (1997). "Modeling extreme events for insurance and finance" Berlin: Springer.

Engle, R.F. (1982). "Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation" Econometrica, Vol. 50, No. 4: pp. 987-1007

Eydeland, A., Wolyniec K., (2003). "Energy and Power Risk Management: New Developments in Modeling, Pricing, and Hedging" John Wiley & Sons, Inc.

Fisher, R. A., Tippett, L. H. C. (1928). "Limiting forms of the frequency distribution of the largest or smallest member of a sample", Proceeding of Cambridge Philosophical Society, 24, 180-190

Gencay, R., Selcuk, F., (2004). "Extreme value theory and Value-at- Risk: Relative performance in emerging markets" International Journal of Forecasting, Volume 20: pp. 287-303

Hill, B. M. (1975), "A simple general approach to inference about the tail of a distribution", Annals of Statistics, 3, 1163-1174

Hoare, J. (1998), "The UK Electricity Market", Managing Energy Price Risk, Risk Publications, pp. 213

Jorion, P. (1997). "Value-at-risk: The new benchmark for controlling market risk" Chicago: McGraw-Hill.

Masson, G. S. (1999). "Competitive Electricity Markets Around the World: Approaches to Price Risk Management" Managing Energy Price Risk, 2nd edition, Risk Books

McNeil, A. J. (1997). "Estimating the tails of loss severity distributions using extreme value theory" ASTIN Bulletin, 27: pp. 1117-1137

McNeil, A. J. (1999). "Extreme Value Theory for Risk Managers", http://www.math.ethz.ch/~mcneil/ftp/cad.pdf

McNeil, A. J., & Frey, R. (2000). "Estimation of tail-related risk measures for heteroscedastic financial time series: An extreme value approach" Journal of Empirical Finance, 7: pp. 271-300

Pickands, J. (1975) "Statistical inference using extreme order statistics" The Annals of Statistics, Vol. 3, No.1, 199-131

Pindyck, R.S & Rubinfeld, D.L. (1998). "Econometric models and economic forecasts", 4th edition, Ivwin/McGraw-Hill

R Project software, http://www.r-project.org

Reiss, R., & Thomas, M. (2001). "Statistical analysis of extreme values", Basel: Birkhauser.

Stock, J. H, & Watson, M.W. (2003). "Introduction to Econometrics", Pearson Education, Inc