UTRECHT UNIVERSITY
DEPARTMENT OF MATHEMATICS

IN PARTIAL FULFILMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
MATHEMATICAL SCIENCES

STATISTICAL ANALYSIS AND FORECASTING OF KLM ROYAL DUTCH AIRLINES GROUND HANDLING PROCESS DURATIONS

Author:
Evrim AKAR

Supervisors
Dr. Eduard BELITSER
Department of Mathematics,
UU
Marc PAELINCK
Department of Decision -
Support, KLM

May 31, 2009

Abstract

It is crucial for organizations of all sizes to make future forecasts in order to decrease the uncertainty of the environment and to take the advantage of opportunities available to the organization. As each data set has its specific properties, not every method results in accurate forecasts for a given data set. Therefore, a variety of techniques have been developed for effective and efficient predictions.

In this study, the analysis of random processes and time-series are examined. Moreover, a brief description about the major forecasting methods are given and a nonparametric curve estimation method is introduced for future predictions. In order to increase the forecast accuracy, ”comparison and combination method” is applied. We focus on the quantile values of Royal Dutch Airlines (KLM) Ground Handling Process Times Data.

Contents

List of Tables

List of Figures

Chapter 1

INTRODUCTION

In order to manage the processes on the day of operations, Royal Dutch Airlines (KLM) has split up the aircraft processes into so called Building Blocks. Each Building Block may consist of various sub-processes, but has one Process Owner with whom the Operations Control Department has Service Level Agreement (SLA) about quality and throughput times.

As a part of the SLA, each Process Owner delivers a forecast of the statistical distribution for their Building Block processes for the coming schedule season. The distributions are described by means of a sequance of percentiles. These statistics are used by the schedule developers and by Operations Control to determine whether a new schedule can meet the punctuality standards which are set by the KLM Management. Moreover, during an operational schedule period, a check is continuously performed to determine whether the actual process times still match the forecasted statistics. Corrective measures may be taken if this is not the case.

The Ground Services Department is one of the Process Owners, and delivers statistical distributions for the ground handling process at Amsterdam Schiphol Airport, the Netherlands, and at all outstations. The statistics consist of a set of percentiles based on historical data.

Until recently, a Gamma distribution was fitted on the given percentiles before performing a simulation with the OpiuM (The performance model OPiuM is a computer simulation model that is used by KLM to evaluate the performance and robustness of a proposed flight schedule) model. Because this introduces an undesired interpretation of the process statistics, it has been decided to replace the Gamma distribution by a distribution consisting of line segments connecting the percentiles. However, the distribution consisting of line segments appears not to give realistic results when a schedule forecast is performed. It is suspected that this is due to inaccuracies in the given statistics. A possible explanation is that some categories have an insufficient number of observations from which to derive the statistics.

This work will consist in developing a statistical model to provide a more accurate forecast of the process statistics, based on historical data.

The observations have been made on different type of aircrafts on different routes on operation plan periods since 2006. Each aircraft differs in size; therefore, the standard ground handling process time changes for each group. Furthermore, the process times depend on the type of the flight, which can be departure, turn-around, or so on.

Table 1.1 shows the aircraft types we studied on and their capacities.

The form of the data we used is shown in Table 1.2.

In table 1.2, *BB24_pln* states the standard ground handling process times which is determined by KLM Management. *"BB24_Act"* is for actual observed process time. *"Date_Ams"* shows the date the flight occurs. The *"Flight"* cell contains the flight number. The departure station of the flight can be found in the *"From"* part and *"To"* cell stands for the destination of the flight. *"KofArr"* and *"KofDep"* denote the region of provenance of the flight preceding the ground handling process and the region of the destination of the flight following the ground handling process, respectively. "*E*" denotes a flight within Europe, "*I*" denotes an intercontinental flight and "*A*" stands for any region. For stations other than Amster-

Aircraft Codes (IATA Codes)	Manifacturer & Aircraft Type	Seating Capacity (Economy+Business)
332	Airbus 330 (200)	221+30
733	Boeing 737-300	88+39 / 112+15
734	Boeing 737-400	108+39 / 114+33
738	Boeing 737-800	114+57
739	Boeing 737-900	132+57
744	Boeing 747-400	378+42
74E	Boeing 747-400 (Combi)	238+42
73H	Boeing 737-800 (Winglets)	114+57
763	Boeing 767-300	245+24
772	Boeing 777-200(ER)	292+35
M11	McDonell Douglas MD-11	270+24
73J	Boeing 737-900 (Winglets)	132+57
73W	Boeing 737-700 (Winglets)	112+12
77W	Boeing 737-300 (Winglets)	390+35

Table 1.1: Aircraft Properties

dam Schiphol Airport, these indicators are not relevant and have the value "*X*". *"Type"* cell is used to record the type of the flight which is one of the following: "*ARR*" (arrival handling - which in this study we are not interested in), "*DEP*" (departure handling), "*T/A*" (Turnaround, which is a combination of arrival and departure handling), "*TR*", "*TRE*", "*TRN*", "*TRS*", "*TRW*" (transit flights on outstations where "*E*" stands for Eastbound, "*N*" for Northbound, "*S*" for Southbound and "*W*" for Westbound). Operational Plan Year and the Operational Plan Period are shown in the cells *"OPyear"* and *"OPseq"*, respectively. *"Regio"* stands for the region of the destination of the flight, for example, looking at the above table, OTP (Bucharest - Henri Coanda International Airport, Romania) is in EUR while JFK (New York - John F. Kennedy International Airport, USA) in ICA. *"Subtype"* shows the type of the aircraft and *"Time_Ams"* is the time in Amsterdam, the

BB24_Pln	1:23	1:23	1:53	0:46
BB24_Act	1:13	1:27	1:52	0:50
Date_Ams	12-Aug-07	28-Jan-07	6-Apr-08	15-Apr-08
Flight	KL 0871	KL 0641	KL 0561	KL 1359
From	AMS	AMS	AMS	AMS
To	DEL	JFK	EBB	OTP
KofArr	A	A	I	A
KofDep	I	I	I	E
Type	DEP	DEP	T/A	DEP
OPyear	2007	2006	2008	2008
OPseq	2	4	1	1
Regio	ICA	ICA	ICA	EUR
Subtype	M11	744	332	73H
Time_Ams	9:01	12:34	9:30	11:02

Table 1.2: Form of the Available Data

Netherlands, when the flight occurs.

The Ground Handling is a process which includes many subprocesses, such as, deboarding, cleaning, catering, boarding, loading and unloading processes. Afhandelings Instructies Schiphol (AIS) is a process guidance manual for divisions involved in dispatching airplains from KLM and KLM-standard contracted customer airlines at KLM-HUB Schiphol. For each division (and third parties), processing steps, handling over issues, tasks, and standardization are completely discribed according to policy principles and budget regulations. Goal is to guarantee safety, punctuality and service level to customers.

Figure 1.1 is a part of AIS handbook which shows the standard ground handling procedure and process times for aircraft type Boeing 737-800 (IATA Code: 738) for departure.

5.1.3C. B737-800

■ NORM-GRONDTIJD

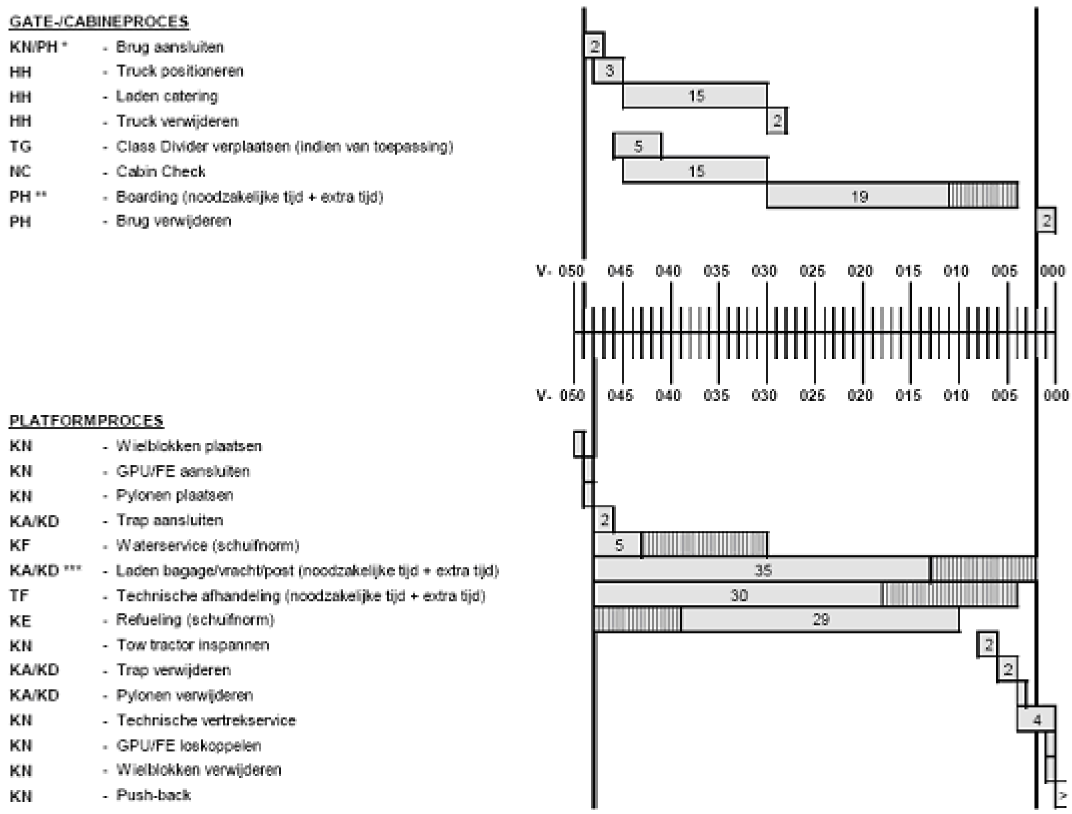
LOS VERTREK	:	50 min
TURNAROUND	:	65 min
KORTE TURNAROUND	:	60 min
LOSSE AANKOMST	:	40 min

■ UITGANGSPUNTEN NORMERING

Zie AIS hoofdstuk 5, paragraaf 5.1.3A. (B737-300)

5.1.3C. B737-800 (vervolg)

■ NORMTIJDEN LOS VERTREK - 50 min (grafische weergave)



* Brug aansluiten: actie KN na sleepbeweging vanaf 06.00 uur.

** Boarding: uitsluitend toegestaan indien cockpitcrew aanwezig in cockpit en gereed voor boarding.

*** Laden ruimen: inclusief openen/sluiten deuren, verwijderen/plaatsen vrachtnet, plaatsen/verwijderen transportbanden en bagagekarren.

■ **NORMTIJDEN LOS VERTREK - 50 min** (chronologische weergave)

V-70/50/45/35 *	PH	GA in gate (10 min vroeger indien geen SA 2)
V-70/40/35/30/25 *	PH	SA 1 in gate (10 min vroeger indien geen SA 2)
V-70/30	PH	SA 2 in gate (optioneel)
V-50	KN	Vliegtuig op positie
V-49	KN/PH	Aanvang brug aansluiten
V-49	KN	Wielblokken geplaatst
V-48	KN	GPU/FE aangesloten
V-48	KN	Pylonen geplaatst
V-48	HH	Aanvang positioneren catering truck
V-48	KA/KD	Aanvang trap aansluiten
V-48	KF	Aanvang waterservice
V-48	KA/KD	TLO / teamleden / materieel bij vliegtuig
V-48	KA/KD	Openen eerste vrachtdeur / aanvang laden ruimen
V-48	TF	Aanvang technische afhandeling
V-47	KN/PH	Brug aangesloten
V-46	KA/KD	Trap aangesloten
V-46	TG	Aanvang verplaatsen Class Devider (indien van toepassing)
V-45	HH	Aanvang laden catering
V-45	NC	Cabincrew in gate
V-45	NC	Aanvang Cabin Check
V-45 – V-30	NC/Cockpit	Uitsluitsel uitgifte cockpit-/crewseats
V-43	KF	Einde waterservice
V-41	TG	Einde verplaatsen Class Devider (indien van toepassing)
V-39	KE	Aanvang refueling
V-30	Cockpit	Cockpitcrew in vliegtuig
V-30	NC	Einde Cabin Check
V-30	PH	Aanvang boarding
V-30/28	HH	Einde laden catering / truck verwijderd
V-30 – V-20	KA/KD	Fysieke informatie-overdracht aan Captain en (Senior) Purser
V-15	PH	Start inventarisatie ontbrekende pax / start gate is closing
V-15 – V-08	PH	Voltooien on-/offloaden pax / informeer TLO over ontbrekende pax
V-14	KL	Afmelden laatste bagage aan KA/KD
V-10	KL	Laatste bagage bij vliegtuig
V-10	KA/KD	Afmelden actuele belading aan KK (indien van toepassing)
V-10	KE	Einde refueling
V-10	PH	Einde ABC boarding
V-10	KA/KD	Contact Captain bij vertraging
V-08	PH	Codeco updated
V-06	KK/KA/KD	Loadsheet in cockpit
V-06	KN	Tow tractor ingespannen
V-06	KA/KD	Aanvang verwijderen trap
V-04	KA/KD	Trap verwijderd
V-04	TF	Einde technische afhandeling
V-04	KN	Aanvang technische vertrekservice
V-04	PH	Einde boardingproces
V-03	KA/KD	Pylonen verwijderd
V-02	KA/KD	Einde laden ruimen / laatste vrachtdeur gesloten
V-02	PH	(Laatste) passagiersdeur gesloten
V-02	PH	Aanvang verwijderen brug
V-00	PH	Brug geparkeerd
V-00	KN	Einde technische vertrekservice
V-00	KN	GPU/FE losgekoppeld
V-00	KN	Wielblokken verwijderd
V-00	KN	Aanvang push-back

* **Meldingsnormtijd GA/SA 1: afhankelijk van aantal geboekte passagiers en/of gate van vertrek**
Vermelde tijden zijn uiterlijke tijden

Figure 1.1: A Part of AIS Handbook

The actual durations of the process are measured by signals which are sent when the aircraft doors are opened and closed.
As stated before the Ground Handling processing time data is assumed to follow a shifted Gamma distribution We, first of all, try to prove if the data comes from a Gamma distribution. We assumed that the data is independent and identically distributed (iid). Moreover, although there are some unrealistic observations, we did not ignore any data as it is not possible to know at this point that the reason of these records is mismeasurement or not, as very long durations may be due to a possible delay and very short durations may belong to the flights with a very few number of passengers.
Assuming the data comes from a shifted Gamma distribution, we first estimated the parameters by using Maximum Likelihood Estimation.
In this paper, we first analyze the statistical properties of ground handling process duration samples. Forecasting procedure of the one period ahead values of the quantiles of the samples is given in Chapter 3. After defining widely used basic forecasting techniques, we try to develop a nonparametric method which gives estimates for τ quantiles for $\tau = 0.1, 0.2, 0.3, \dots, 0.9$. Estimation and improvement of model parameters are also examined in this chapter. Applying the defined methods to given samples, forecast accuracies of the methods are compared. The best suited method is expected to be chosen by combination. (Conclusion has not been made yet.)

Chapter 2

EMPIRICAL ANALYSIS

In this chapter, we examined the statistical properties of Ground Handling Process Duration samples. Among many data sets, we chose a test set from year 2008 - Operation Period 2. This set contains ground handling process times of the *Boeing* 737 – 400 (734) type aircrafts having departed from Amsterdam Schiphol Airport(AMS), landed in Europe. Our set contains 2725 observations which would yield good results for testing purposes (see Appendix A - Maximum Likelihood Estimation).
First, we tried to determine the distribution of the data based on the sub-processes using Maximum Likelihood Estimation. Then, we followed Least Squares Estimation method using the quantiles of the data. In order to test the results, we used Chi-Square Goodness of Fit Test.
For numerical computation purposes the statistical software R has been used.

2.1 DISTRIBUTION ANALYSIS OF THE DATA

In this section, we tried to find a suitable distribution which fits to our data. The sub-processes are distributed according to either Gamma distribution or Weibull distribution ([11]Boom, L[2003]). Therefore, we tested if the whole ground handling process is distributed according to one of those distributions.

We applied defined maximum likelihood process to the data by using statistical software R. Clearly, ground handling process takes time greater than 0. However, the smallest time this process takes is not close to 0 which can be easily explained by practical reasons. Therefore, if the distribution is gamma, there must be a shift modification in our data, as Gamma or Weibull distributed random variables start close to 0. Thus, we used an additional shift parameter for both distributions. First, the Gamma distribution is tested.

We estimate $\theta = (s, k, \lambda)$ in Θ where s is the shift parameter, k is the shape and λ is the scale parameters.

The corresponding R code can be found below:

```
mlogl < -function(theta, x){sum(-dgamma(x - theta[1], shape = theta[2],
+scale = theta[3], log = TRUE))}
theta.start < -c(1, (mean(x)) * *2/var(x), var(x)/mean(x))
out < -nlm(mlogl, theta.start, x = x)
theta.hat < -out$estimate
theta.hat
```

The resulting R output is as follows:

```
theta.hat
[1] 3.170722  19.604990  1.986347

$minimum
[1] 9744.346

$gradient
[1]  - 1.388600e - 05   - 3.699634e - 05   - 4.010511e - 04

$iterations
```

[1]26

Here, "x" is our random vector of length $n = 2725$, "*theta = (theta[1], theta[2], theta[3])*" represents our set $\theta = (s, k, \lambda)$ and "*mlogl*" is the log-likelihood function. We chose "$theta.start = (1, (mean(x))^2/var(x), var(x)/mean(x))$" as starting point because sample shape, k_n, and sample rate, λ_n, are calculated as:

$$k_n = \frac{\mu_n^2}{\sigma_n^2} \quad \text{and} \quad \lambda_n = \frac{\sigma_n^2}{\mu_n}$$

where μ_n is sample mean and σ_n^2 is sample variance. These are the best estimators one can calculate from the sample.

As seen in the R output, the parameters minimizing log-likelihood function are

$$\theta = (s, k, \lambda) = (3.170722, \quad 19.604990, \quad 1.986347).$$

Now, we have to test how good the Gamma distribution with these parameters fits to our data.

First of all, a quantile-quantile graph may draw a representative picture for our aim. $y = x - s$ is used in Figure 2.1 The q-q plot shows that our data does not follow the estimated Gamma distribution. In order to have more theoretical conclusion, Chi-square goodness of fit test is applied. Our null hypothesis is that the samle follows a Gamma distribution with the estimated paameters. Because there is not a general optimal choice for the bin-width, in this study, we use the following approach: First, we divide our data into $2n^{2/5}$ groups. Then, we check the number of observations in each bin and rearrange the bins such that each bin has at least 5 observations.

Total number of observations, n, we have is 2725, so $2n^{2/5}$ is approximately 47. Therefore we divide the interval [1.829278, 149.879278] into 47 equal sub-intervals with the bin-width = 3.15.

The R function "`hist`" is used for this purpose. The number of observations in each bin is:

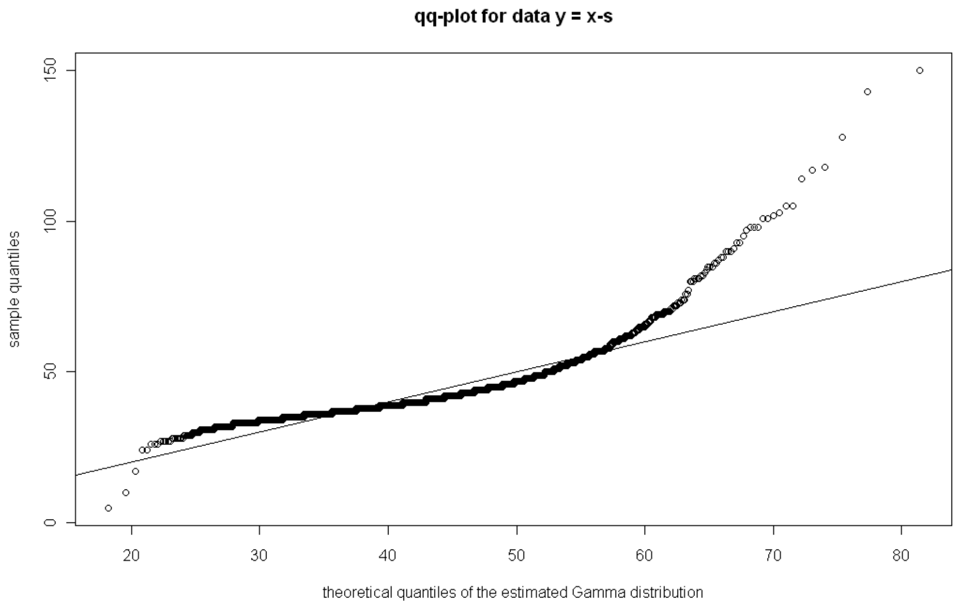


Figure 2.1: Gamma Q-Q Plot for shifted GHP Times

1 1 0 1 0 0 10 25 145 469 674 524 304 220 90 66 50 29 22 22
17 10 3 3 7 6 6 3 1 4 3 3 0 0 1 2 0 0 0 1 0 0 0 1 0 1

As one observes, there are many empty bins as well as the bins which include less than 5 observations. We rearrange the bins as follows:

$(0, 24], (24, 27], (27, 30], (30, 33], (33, 36], (36, 39], (39, 42], (42, 48], (48, 56],$
$(56, 71], (71, 90], (90, 102], (102, 150]$

so that the observations, x_i's, in each bin are:

13 25 145 469 674 524 304 281 151 94 28 11 6

Figure 2.2 and Figure 2.3 show the histograms of the former and the latter cases. The test statistics is, $\chi^2 = \frac{\sum_{i=1}^{k} (O_i - E_i)^2}{E_i} = 105370.4$, which gives a p-value very

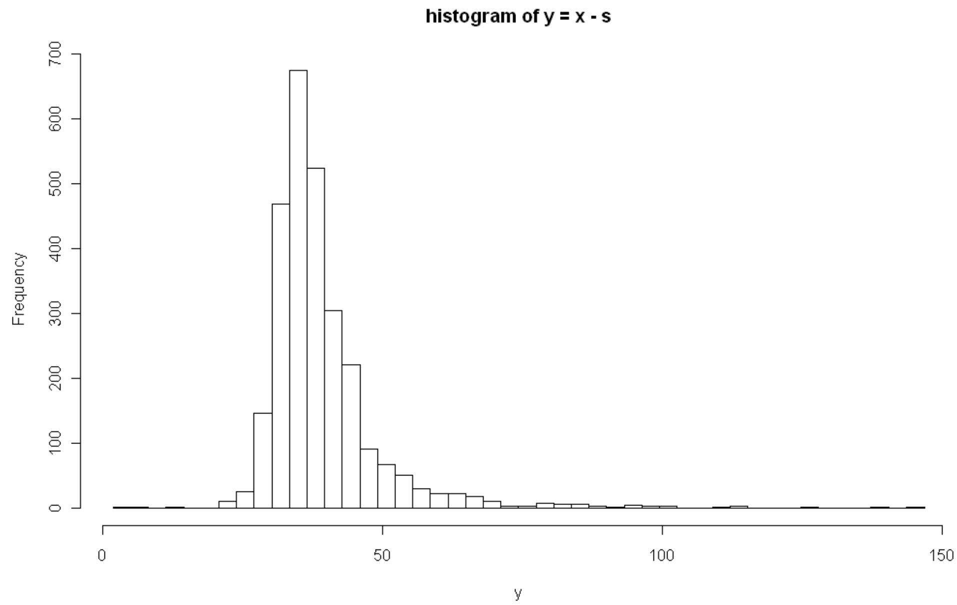


Figure 2.2: Histogram of Shifted GHP Data with bin-width=3.15

close to 0 with $13 - 1 - 2 = 10$ degrees of freedom. Therefore one concludes that we cannot accept the null-hypothesis and Gamma distribution with estimated parameters does not fit to our sample.
Because for this particular study we are interested in predicting the quantiles between 0.1 and 0.9, one may want to see how our data behaves between 10th and 90th percentiles. In order to find out this, first we determine the sample quantiles.

2.2 QUANTILES

Any real-valued random variable X can be characterized by its (right continuous) distribution function

$$F(x) = \mathbb{P}(X \leq x)$$

where the τ-quantile for any $0 < \tau < 1$ is:

$$F^{-1}(\tau) = \inf\{x : F(x) \geq \tau\}$$

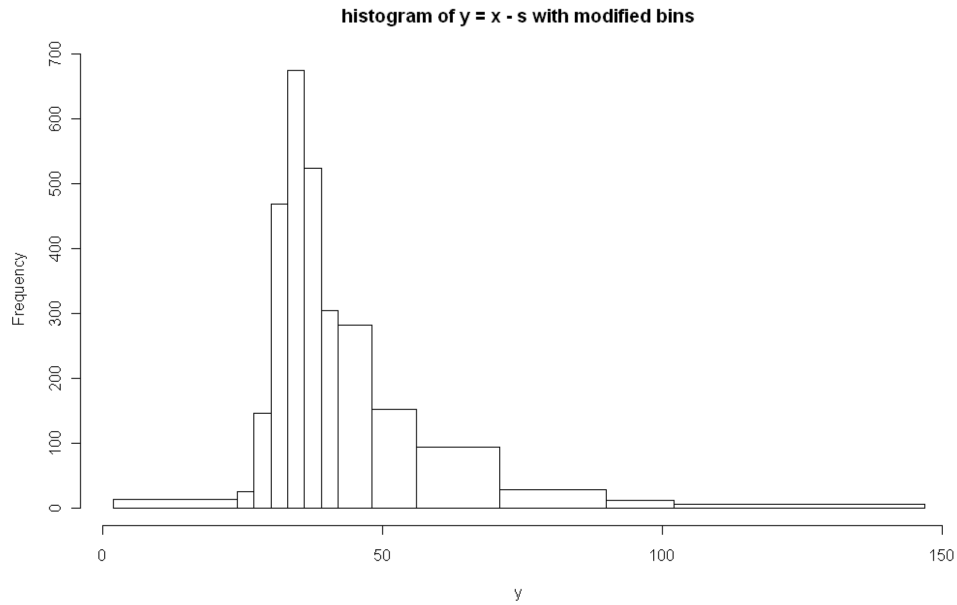


Figure 2.3: Histogram of Shifted GHP Data with Rearranged Bins

A simple optimization problem is used to find the quantiles as follows:
Under the piecewise linear loss function, illustrated in Figure 2.4, $\rho_\tau(u) = u(\tau - 1\{u < 0\})$ for some $\tau \in (0, 1)$, a point estimate is required for a random variable, X, with (posterior) distribution function F. One would like to minimize the expectation of the above loss function, ρ, given X. Fox and Rubin (1964) studied the admissibility of the quantile estimator under this loss function. We would like to minimize the expectation

$$\mathbb{E}\rho_\tau(X - \hat{x}) = (\tau - 1)\int_{-\infty}^{\hat{x}} (x - \hat{x})dF(x) + \tau \int_{\hat{x}}^{\infty} (x - \hat{x})dF(x)$$

Differentiating with respect to $\hat{x}$, we obtain

$$0 = (1 - \tau)\int_{-\infty}^{\hat{x}} dF(x) - \tau \int_{\hat{x}}^{\infty} dF(x) = F(\hat{x}) - \tau$$

Any element of $\{x : F(x) = \tau\}$ minimizes the expected loss because F, being a distribution function, is monotone. The result is either unique (in that case $\hat{x} = F^{-1}(\tau)$)

or an interval of τ quantiles (from which the smallest to be chosen for convention that the empirical distribution function is left continuous).

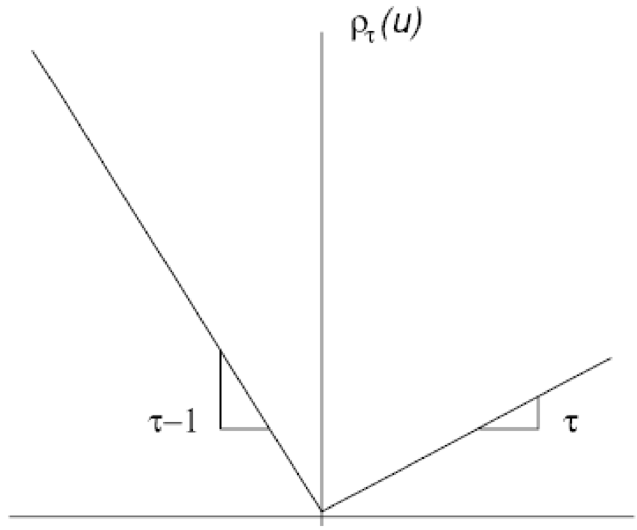


Figure 2.4: Quantile Regression ρ function

If we replace F by the empirical distribution function F_n, where $F_n = n^{-1} \sum_{i=1}^{n} 1\{X_i \leq x\}$, we may still choose $\hat{x}$ to minimize the expected loss:

$$\int \rho_\tau(x - \hat{x}) dF_n(x) = n^{-1} \sum_{i=1}^{n} \rho_\tau(X_i - \hat{x})$$

The resulting $\hat{x}$ would give us the τ-*sample* quantile. Thus, denoting the τ sample quantile by $q(\tau)$, one can formulize finding the quantile as follows:

$$q(\tau) = argmin_{\hat{x}} = n^{-1} \sum_{i=1}^{n} (X_i - \hat{x})(\tau - 1\{X_i - \hat{x} < 0\})$$

We now have expressed the τ-sample quantile as a linear optimization problem which can be solved with well known methods such as interior point and simplex algorithms [9]Koenker [2005].

Royal Dutch Airlines (KLM) is interested in the quantiles of the ground handling process times in order to test and improve the quality of the schedules. Historical data are available for 3 years, since 2006, each year divided into 4 operation plan (OP) periods.We would like to know more about the τ quantile for the next OP period based on the historical τ quantiles.

2.3 ESTIMATION OF THE DISTRIBUTION : AN OPTIMIZATION APPROACH

In this section, we would like to use an optimization approach in order to determine the distribution of the data. In the previous section, we used Maximum Likelhood Estimator, however the estimated Gamma distribution did not fit to our whole sample. Unlike the previous approach, here, we try to see if a Gamma distribution can be fit to our data when we cencor the observations between 10th and 90th percentiles. In other words, let a and b denote the 0.1 and 0.9 quantiles, respectively. We use the data X_i, where $a \leq X_i \leq b$ to see how good an estimated Gamma distribution fits to the cencored part of the sample. We construct a minimization problem and use the statistical software `R` to solve it.

Lets first look at the τ sample quantiles where $\tau \in \{0.1, 0.2, \ldots, 0.9\}$. Using "`kuantile`" function of `R`, which calculates the sample quantiles according to the above rule, we obtained:

```
>kuantile(x, c(0.1, 0.9))
>percentiles: 10% 90%
>quantiles:    34   52
```

In Figure 2.5, one can see the graph of sample quantiles: Now we would like to find the parameters of the Gamma distribution whose quantiles between 10% and

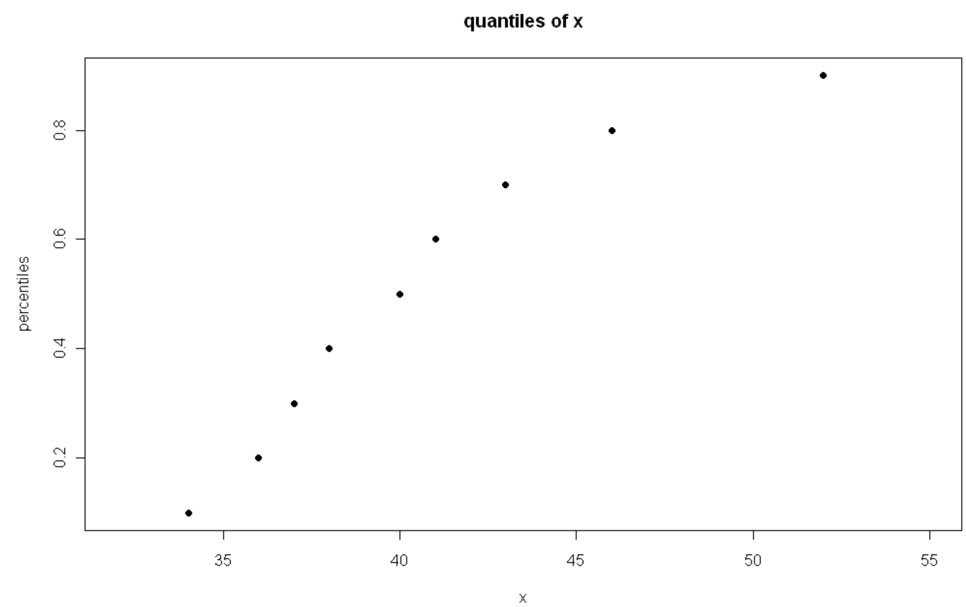


Figure 2.5: Quantiles of GHP Durations of Year 2008 OP2

90% are closest to our sample. In order to find these, we tried to minimize the sum of squares of the horizontal distance between the sample quantiles and theoretical Gamma quantiles. In this case, we cannot use the usual maximum likelihood algorithm as we censored our sample in the interval between 0.1-quantile and 0.9-quantile. Since this part of the sample we are using does not give the whole set of observations, the distribution parameters one finds using the usual MLE algorithm would give high errors between the estimated quantiles and the actual sample quantiles. We define the sum of squares of the distance between sample quantiles and theoretical quantiles as

$$d = \sum_{\tau} (\mu_\tau - s - F^{-1}(\tau, k, \lambda))^2$$

where μ_τ is the τ sample quantile, τ in (0,1) (in our case, we are interested in $\tau \in \{0.1, 0.2, 0.3, \ldots, 0.9\}$) and F is the theoretical cumulative Gamma distribution function.

We would like to minimize the above quantity over s, k and λ, which are shift, shape and scale parameters, respectively. The statement of the problem is:
Given μ_τ

$$\min_{s,k,\lambda} d = \sum_\tau (\mu_\tau - s - F^{-1}(\tau, k, \lambda))^2$$

subject to:

$$s \geq 0, k \geq 0, \lambda \geq 0$$

F cumulative Gamma distribution function.

We used the statistical software R for this simple optimization problem. You can find the corresponding code below:

```
>p<-function(x) kuantile(x,seq(0.1,0.9,0.1))
>i<-seq(0.1,0.9,0.1)
>fh<-function(theta,x)sum((p(x)-theta[1]-qgamma(i,shape=theta[2],
>+scale=theta[3],log=F))**2)
>theta.start<- c(var(x)/mean(x) ,(mean(x))**2/var(x),1)
>out<-function(x)nlm(fh,theta.start,x=x)
```

In the above code, `fh` is the previously defined d function. `qgamma` is an R function which finds the quantiles for given probability, shape and scale for Gamma distribution. We used "`theta`" for our unknown parameters s, k and λ, and minimize the function `fh` over θ using R minimization function `nlm`. `theta.start` defines the starting point for the minimization algorithm and we calculated it from the sample as before.
We used here the data set of departure flights in 2008 Operation Plan Period 2, aircraft type 734, departure station Amsterdam Schiphol Airport, and arrival station in Europe.
Here is the R output of the above defined code:

```
minimum value of the function "out"
```

```
$minimum
[1]  1.380391

the estimated Gamma parameter minimizing the defined distance
$estimate
[1]  33.350446   1.192140   6.942234

number of iterations
$iterations
[1]  100
```

Figure 2.6 shows the sample quantiles and the Gamma distribution curve which is found by minimizing the sum of squares of the distance between quantiles.

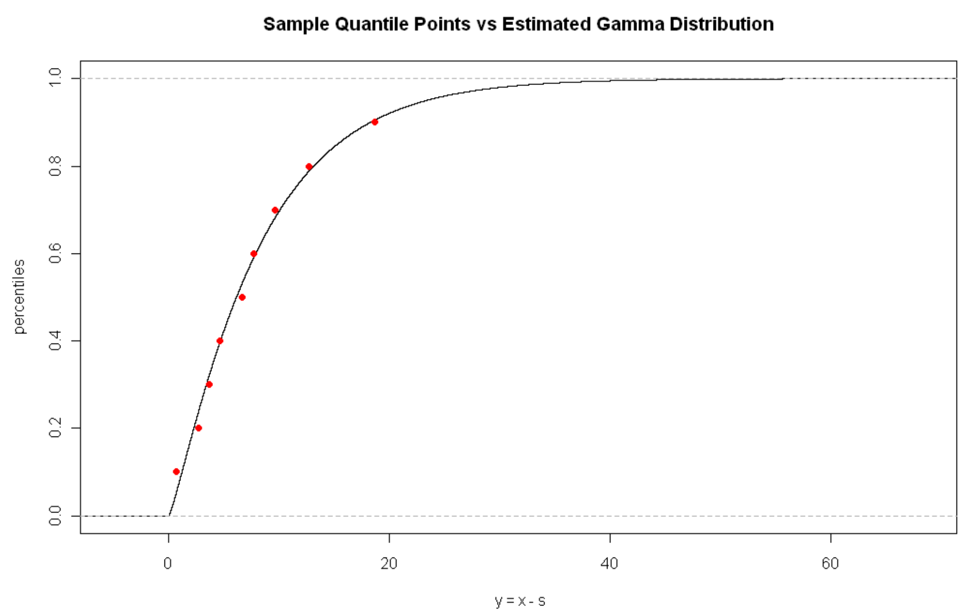


Figure 2.6: Gamma Distribution Fit by Optimizing the Distance between Theoretical and Sample Quantiles

We also tested if the distribution is Gamma for the censored set of observations censored at 0.10- and 0.90- quantiles. First, let's look at the quantile - quantile plot of the sample quantiles and the theoretical estimated Gamma quantiles shown in Figure 2.7. As one observes from the graph, the sample data is rounded and

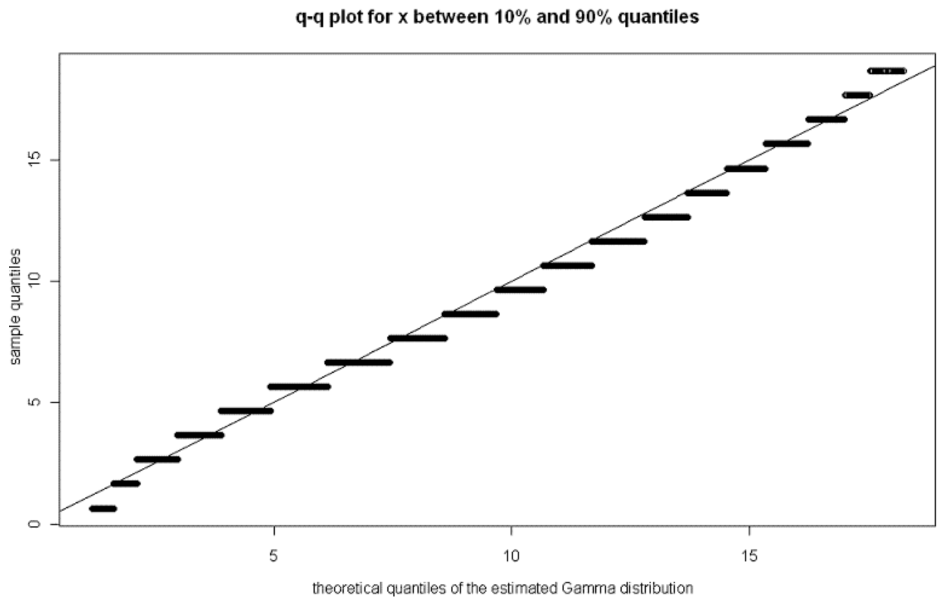


Figure 2.7: Quantile-Quantile Plot of the Shifted Sample Quantiles and Estimated Gamma Quantiles

grouped; therefore, the number of unique points in the sample is too few. However, it is still difficult to conclude if the sample distribution is Gamma when the outliers are censored. Therefore, we applied χ^2 Goodness of Fit Test with

H_0 = The censored part of the sample comes from a Gamma distribution with the estimated parameters.

H_1 = The distribution of the censored part of the sample is not Gamma with the parameters.

Applying the given χ^2 formula

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} = 226.4756$$

The corresponding p value is very close to 0 with 7 degrees of freedom. Therefore, we have to reject the null hypothesis saying the censored part of the data between 0.1 and 0.9 quantiles are disttibuted according to Gamma distribution.

2.4 DISTRIBUTION ANALYSIS: WEIBULL DISTRIBUTION

After we conclude the distribution of the sample is not a Gamma distribution, in this section we tested if the sample comes from a Weibull Distribution.
Remember that the probability density function of 2-parameter Weibull distribution with shape k and scale λ is

$$f(x) = \frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1} e^{-(x/\lambda)^k}, \quad x \geq 0, \ \ k > 0, \ \ \lambda > 0$$

The corresponding cumulative distribution function is

$$F(x) = 1 - e^{-(x/\lambda)^k}$$

As stated in Seciton $A.3$ the MLE estimators, $\hat{k}, \hat{\lambda}$ of the shape and the scale,respectively, are

$$\frac{1}{k} = \frac{\sum_{i=1}^{n} x_i^k \log x_i}{\sum_{i=1}^{n} x_i^k} - \frac{1}{n}\sum_{i=1}^{n} \log x_i \qquad (2.1)$$

$$\hat{\lambda} = \left(\frac{\sum_{i=1}^{n} x_i^k}{n}\right)^{1/k} \qquad (2.2)$$

Applying the above formulas in the statistical software R, one gets

$$\hat{k} = 3.445772 \qquad (2.3)$$

$$\hat{\lambda} = 46.0154 \qquad (2.4)$$

Lets first look at the q-q plot of the sample quantiles and the estimated Weibull quantiles. Figure 2.8 shows clearly that the distribution of the sample is not Weibull. Next, we apply the optimization approach with Weibull Distribution. In this case,

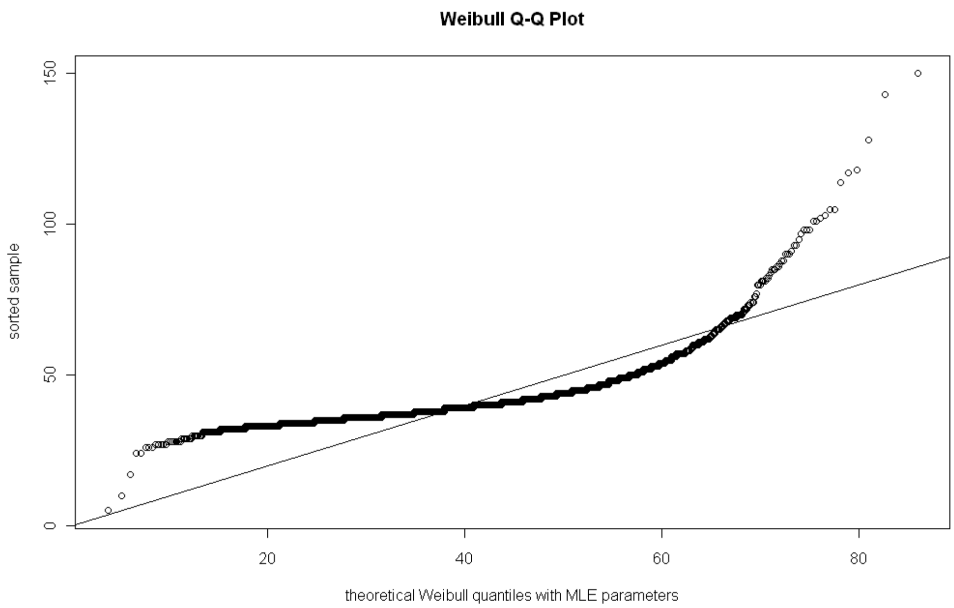


Figure 2.8: Quantile-Quantile Plot of the Theoretical Weibull Quantiles with MLE parameters and the Sample Quantiles

the minimization problem becomes
For $\tau \in \{0.1, 0.2, \ldots, 0.9\}$
Given τ quantiles μ_τ

$$\min_{s,k,\lambda} d = \sum_{\tau} (\mu_\tau - s - F^{-1}(\tau, k, \lambda))^2$$

subject to:

$$s \geq 0, k \geq 0, \lambda \geq 0$$

F cumulative Weibull distribution function.

We used the statistical software R for this simple optimization problem. You can find the corresponding R code below:

```
>p<-function(x) kuantile(x,seq(0.1,0.9,0.1))
>i<-seq(0.1,0.9,0.1)
>fh<-function(theta,x)sum((p(x)-theta[1]-qweibull(i,shape=theta[2],
>+scale=theta[3],log=F))**2)
>theta.start<- c(1,1,1)
>out<-function(x)nlm(fh,theta.start,x=x)
```

The above code is the same as defined for Gamma distribution. We only change R function `qgamma` to `qweibull` to find the theoretical Weibull distribution quantiles. Figure 2.9 shows the fit. Now, one would like to see how good this fit is; therefore,

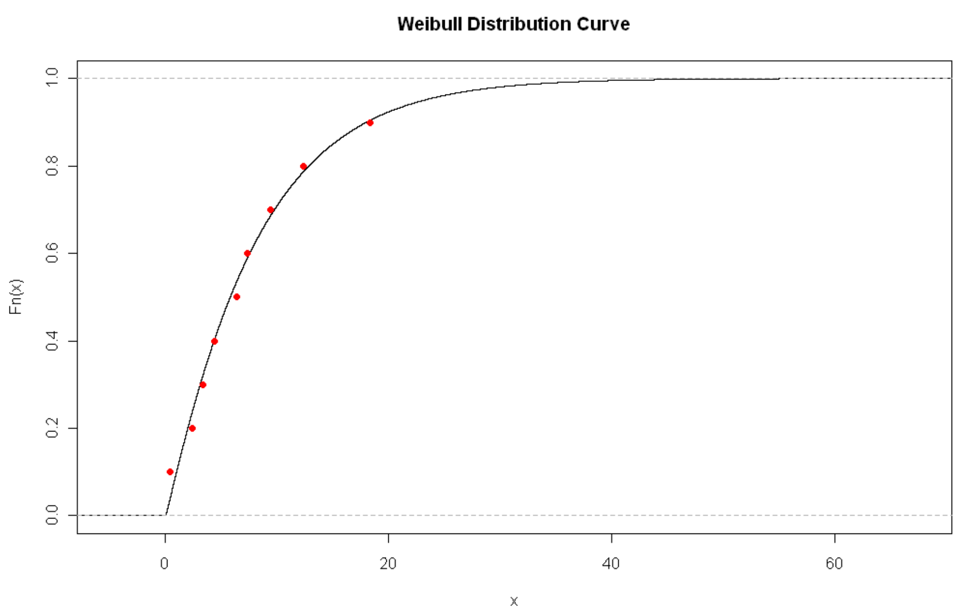


Figure 2.9: Weibull Distribution Fit by Optimizing the Distance between Theoretical and Sample Quantiles

we applied χ^2 Goodness of Fit Test with
H_0 = The censored part of the sample comes from a Weibull distribution with the estimated parameters.
H_1 = The distribution of the censored part of the sample is not Weibull with the estimated parameters.
Applying the defined χ^2 formula

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} = 271.4117$$

The corresponding p value is very close to 0 with 7 degrees of freedom. Hence we conclude that the censored observations do not follow Weibull distribution.
Since the determination of the distribution of the ground handling process is not straightforward, we suspect it has a mixed distribution. Examining the distribution and afterwords conducting the forecasting process are time consuming. Instead, one can predict each quantile for the next operation plan period using the historical sample quantiles. For this procedure, there are many methods, parametric and non-parametric, such as linear, quantile, non-parametric regression or exponential smoothing. In this paper, we first analyse the properties of the quantile time-series we have, then describe some of the prediction methods above mentioned and at the end develop a robust recursive nonparametric curve estimatior in order to estimate the regression function.

Chapter 3

FORECASTING

Many business and economics data are non-stationary and may contain trend and/or seasonal variations. Periodic and recurrent pattern seasonality may be caused by weather, holidays, or repeating promotions. A stochastic trend is often accompanied with the seasonal variations and can have a significant impact on various forecasting methods. In this section, we will compare some forecasting methods for time-series with and without seasonal and/or trend patterns. These methods are Regression, Exponential Smoothing Algorithms and Robust Recursive Nonparametric Curve Estimation. We aim to find the effects of seasonality on our data and we study on the efficiency of the forecasting performance, in particular to answer if a forecasting method concerning trend and/or seasonality yields better forecasts for the ground handling process time data.
We are interested in forecasting the quantiles of ground handling process time data; therefore, our series will be constructed by 10% to 90% quantiles of the data from year 2006 to 2008. Each year is divided into 4 operation plan periods, so we have 12 observations for each quantile.

First of all, in order to see if there is a trend/season effect on our data, let's examine the quantile behaviours. The next section explains the methods to analyze a time series and applications.

3.1 METHODOLOGICAL TOOLS FOR ANALYZING TIME-SERIES DATA

Although our observations do not have time-series characteristics, if we calculate the τ quantiles, $\tau \in \{0.1, 0.2, \ldots, 0.9\}$, for each operation plan period for the same subgroup, we have time-series data constructed by the τ quantiles of our samples. There are certain methods which are used to determine the statistical properties of time-series and therefore gives an insight about the appropriate model that can be applied to the data. In this section, we will concentrate on those methods. Denote the time-series by $Y = Y_t = \{Y_1, Y_2, \ldots, Y_n\}$.

3.1.1 Plotting the Data

A graphical plot of the data is a classical starting point which gives the first insight about the data. It visually helps to identify the existence of the trends and the seasonality.

3.1.2 The Autocorrelation Coefficient

The auto-correlation coefficient is the key statistics in the time-series analysis which indicates the relation between the values in the time-series Y.

The first-order autocorrelation coefficient is the simple correlation coefficient of the first $n-1$ observations, Y_t, $t = 1, 2, \ldots, n-1$ and tehe next $n-1$ observations, Y_t, $t = 2, 3, \ldots, n$. The correlation between Y_t and Y_{t+1} is given by

$$r_1 = \frac{\sum_{t=1}^{n-1} (Y_t - \bar{Y}_{(1)})(Y_{t+1}\bar{Y}_{(2)})}{\sqrt{\sum_{t=1}^{n-1} (Y_t - \bar{Y_{(1)}})^2} \sqrt{\sum_{t=1}^{n-1} (Y_t - \bar{Y}_{(2)})^2}}$$

where $\bar{Y}_{(1)}$ is the mean of the first $n-1$ observations and $\bar{Y}_{(2)}$ is the mean of last $n-1$ observations. For sufficiently large n, assuming the stationarity in the mean, one can simplify and generalize the formula, for $1, 2, \ldots, k$ time lags, as follows:

$$r_k = \frac{\sum_{t=1}^{n-k} (Y_t - \bar{Y})(Y_{t+k} - \bar{Y})}{\sum_{t=1}^{n} (Y_t - \bar{Y})^2}$$

where $r_k = r_{Y_t Y_{t-k}}$, and $\bar{Y} = \sum_{t=1}^{n} Y_t$ is the overall mean.The plot of the autocorrelation function as a function of lag is also called the *correlogram*.

In order to investigate the properties of an empirical time-series, the autocorrelation coefficient is a useful tool. In theory, if we assume an infinite sample, all autocorrelation coefficients of a series with random numbers should be 0. In practise, this situation is not observed. However, we can determine the properties of the distribution of the autocorrelation coefficients of a series.

As Anderson (1942), Bartlett (1946), and others showed, the autocorrelations of random data have a sampling distribution that can be approximated by a normal curve with mean 0 and standard error $1/\sqrt{n}$. One can use this information to develop tests similar to *F-test* or *t-test* and determine the confidence intervals for which the values of the series are random. For this purpose, the area under the normal curve is used.

3.1.3 The Periodogram and Spectral Analysis

Another way to analyse a time-series is to decompose it into a set of sine waves of different frequences. This method can be useful to identify the randomness and seasonality in the time-series.

In a discrete time-series, because there are noangles to deal with, one uses the time units instead of wavelength. As stated in Makridakis, Wheelwright and McGee (1983), any time-series, composed of n equally spaced observations, can be decomposed by least-squares fitting into a number of sine waves of given frequency, amplitude, and phase, subject to the following conditions:

- if n is an odd number, a maximum of $(n-1)/2$ sine waves can be fitted.
- if n is an even number, a maximum of $(n-2)/2$ sine waves can be fitted.

This method was originally known as periodogram analysis (Schuster, 1898) and is variously known as harmonic analysis, Fourier analysis, or spectarl analysis. It helps to identify

- randomness in the data,

- seasonality in a time series,
- the predominance of positive or negative autocorrelation (nasil yani? and eee).

3.1.4 The Partial Autocorrelation Coefficient

The partial autocorrelation at lag k is the autocorrelation between Y_t and Y_{t-k} that is not accounted for by lags 1 through $k-1$.
Partial autocorrelations are used to measure the degree of association between Y_t and Y_{t-k}, when the effects of other time lags - $1, 2, \ldots, k-1$ - are somehow partialled out. Their purpose of use, specifically, to identify the order of an autoregressive model. The sample partial autocorrelation plot is examined to decide on the order of the model if the sample autocorrelation plot indicates that an autoregressive model may be appropriate.
Denote the k-th partial correlation by α_k in autoregressive model:

$$AR(p) : \mu(\epsilon_t|\epsilon_1, \ldots, \epsilon_{t-1}) = \alpha_1\epsilon_{t-1} + \alpha_2\epsilon_{t-2} + \alpha_p\epsilon_{t-p}$$

Define the null hypothesis, $H_0 : \alpha_k = 0$. Then, if r_k is the k-th partial autocorrelation coefficient, then $SE(r_k) \cong 1/\sqrt{n}$ and r_k is approximately normally distributed. The approximate 95% confidence interval for the partial autocorrelations are at $(+/-)2/\sqrt{n}$. Therefore, we reject H_0 if $|r_k| > 2/\sqrt{n}$. If there are p significant partial autocorrelations, then the order should be $AR(p)$.
To sum up, if the process is autoregressive, i.e, its autocorrelation coefficients decline to 0 exponentially, the partial autocorrelations can be examined to determine the order of the process. That order is equal to the number of significant partial autocorrelations.

3.1.5 Examining Stationarity and Seasonality in a Time-Series

The term *stationarity* is used when there is no significant growth or decline in the data, i.e, the data vary around a constant mean, independent of time, and the variance of the fluctuation remains essentially constant over time.

The stationarity or the nonstationarity is often recognized by using the plot of the row data. The autocorrelation plot can also show the nonstationarity quite readily. The autocorrelation coefficients of nonstationary data are significantly different from 0 for several time periods, while the ones for a stationary series drop to 0 after the second or the third time lag.
Positively correlated succesive values indicate that there is a trend (linear or nonlinear) in the data.
Seasonality is defined as a pattern that repeats itself over fixed intervals of time. If the pattern is consistent over the fixed intervals of length i, the autocorrelation coefficient of length i lags will have a high positive value indicating the existence of seasonality. If the autocorrelation coeffient is not significantly different from 0, one concludes that there is no consistent pattern one interval of length i to the other, which indicates the intervals length i apart are unrelated.

3.2 APPLICATIONS WITH KLM GROUND HANDLING PROCESS TIMES DATA

In this section, we applied the above mentioned methods to KLM ground handling process data. First we determined the τ quantiles, $\tau \in \{0.1, 0.2, \ldots, 0.9\}$. For this purpose, we used 2 different data sets, one is for Boeing 737-400 type aircrafts used for intercontinental departure flights from Amsterdam Schiphol Airport and the other is for Boeing 737-400 type aircrafts used for Europe flights again from the same airport.
First, we plotted the behaviours of the τ quantiles over operation plan periods. Figure 3.1 and Figure 3.2 belong to the former group with $\tau = 0.1$ and $\tau = 0.2$, respectively.
Due to few number of data available, it is difficult to identify if there is an trend and/or seasonality from the plots of raw data.
Next, we computed the autocorrelation coefficients.

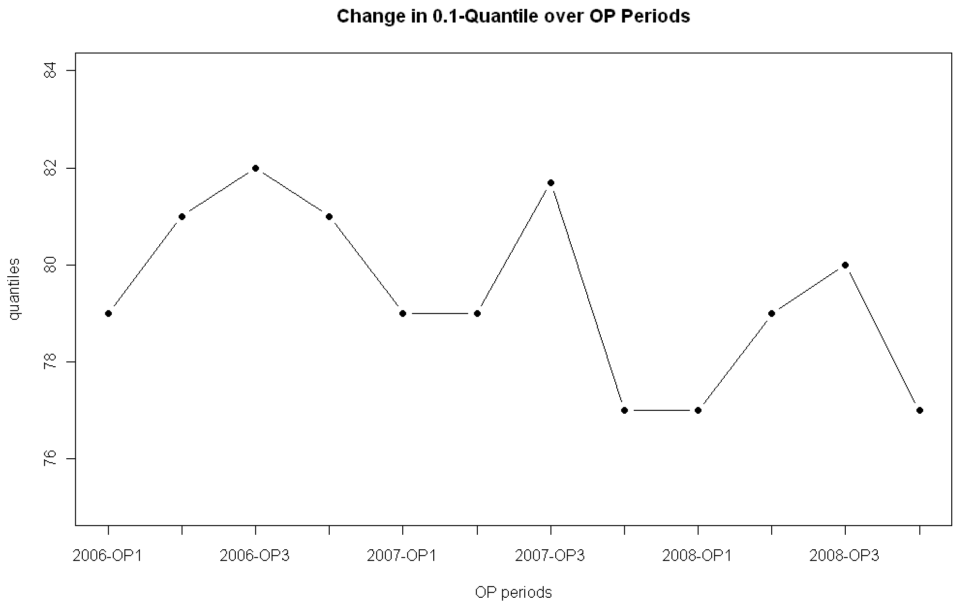


Figure 3.1: Graphical Representation of The 0.1 Quantiles of GHP Time of Boeing 737-400(ICA) over OP Periods

The following analysis are conducted for the first group. Observe that for $\tau = 0.1$, we have the time-series:

$$Y_t = \{79, 81, 82, 81, 79, 79, 81.7, 77, 77, 79, 80, 77\}$$

Hence,

$$Y_{t-1} = \{81, 82, 81, 79, 79, 81.7, 77, 77, 79, 80, 77\},$$

$$Y_{t-2} = \{82, 81, 79, 79, 81.7, 77, 77, 79, 80, 77\},$$

$$Y_{t-3} = \{81, 79, 79, 81.7, 77, 77, 79, 80, 77\},$$

$$\ldots$$

First we compute the means, μ_{t-k} for Y_{t-k}, $k = 1, 2, \ldots, n$.

$$\mu_t = 79.39167$$

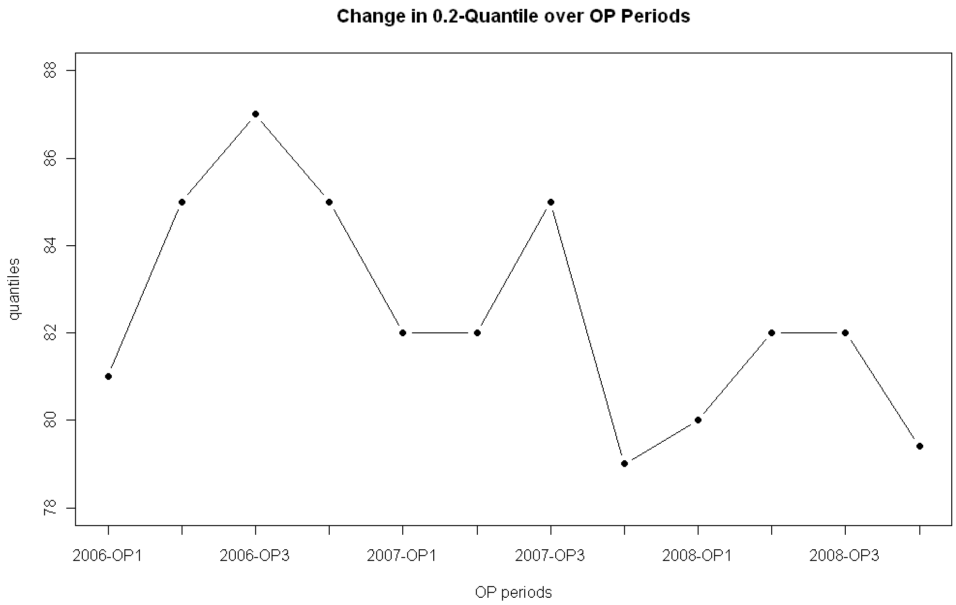


Figure 3.2: Graphical Representation of The 0.2 Quantiles of GHP Time of Boeing 737-400(ICA) over OP Periods

$$\mu_{t-1} = 79.42727$$
$$\mu_{t-2} = 79.27$$
$$\mu_{t-3} = 78.9667$$
$$\ldots$$

Remember the autocorrelation coeeficient formula for $1, 2, \ldots, k$ time lags:

$$r_k = \frac{\sum_{t=1}^{n-k} (Y_t - \bar{Y})(Y_{t+k} - \bar{Y})}{\sum_{t=1}^{n} (Y_t - \bar{Y})^2}$$

Below we apply this formula to Y for $k = 1$ and $k = 2$, respectively:

$$r_1 = \frac{\sum_{t=1}^{n-1} (Y_t - \bar{Y})(Y_{t+1} - \bar{Y})}{\sum_{t=1}^{n} (Y_t - \bar{Y})^2} = 0.1642473$$

$$r_2 = \frac{\sum_{t=1}^{n-2} (Y_t - \bar{Y})(Y_{t+2} - \bar{Y})}{\sum_{t=1}^{n} (Y_t - \bar{Y})^2} = -0.1454469$$

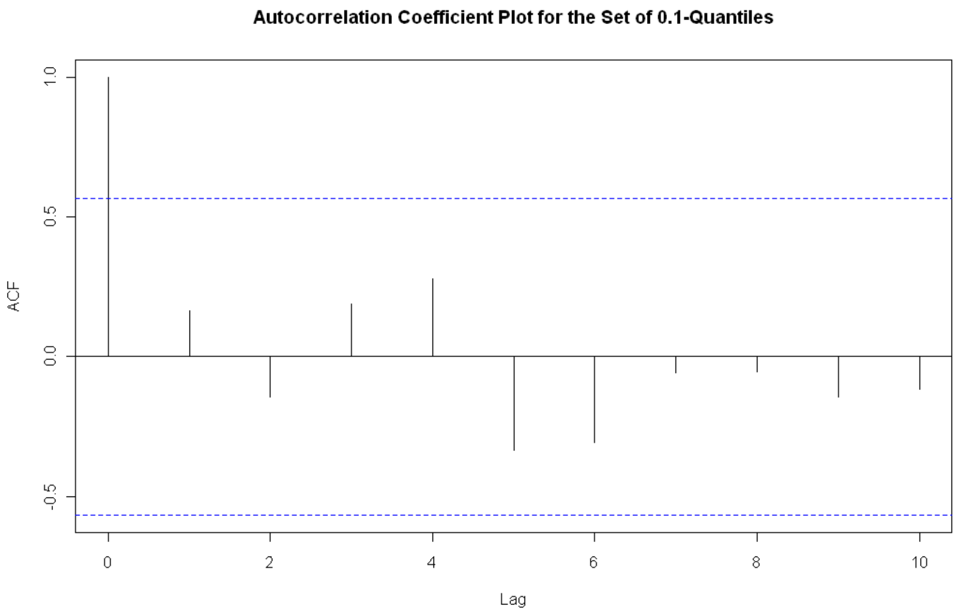


Figure 3.3: Autoregression Coefficients Graph for 0.1 Quantiles of GHP Times over Lags

The autocorrelation coefficient graph with respect to number of lags is given in the Figure 3.3

In order to have a general insight about the behaviours of the τ-quantiles for different τ's, we found the autocorrelation coeffiecient of the series constructed by different τ-quantiles over OP periods. The autocorrelation coefficient graph of 0.5 quantiles (median) is given by the Figure 3.4:

The reader can find the autocorrelation coefficient graphs for all quantiles in the Appendix B.

We followed the same procedure above with the latter series. The results can be found in Appendix B.

After finding the autocorrelation coefficients and plotting them, the first question that should be asked is whether the series is random or not. As explained before, for a random series, lagged values of the series are uncorrelated and r_k is expected

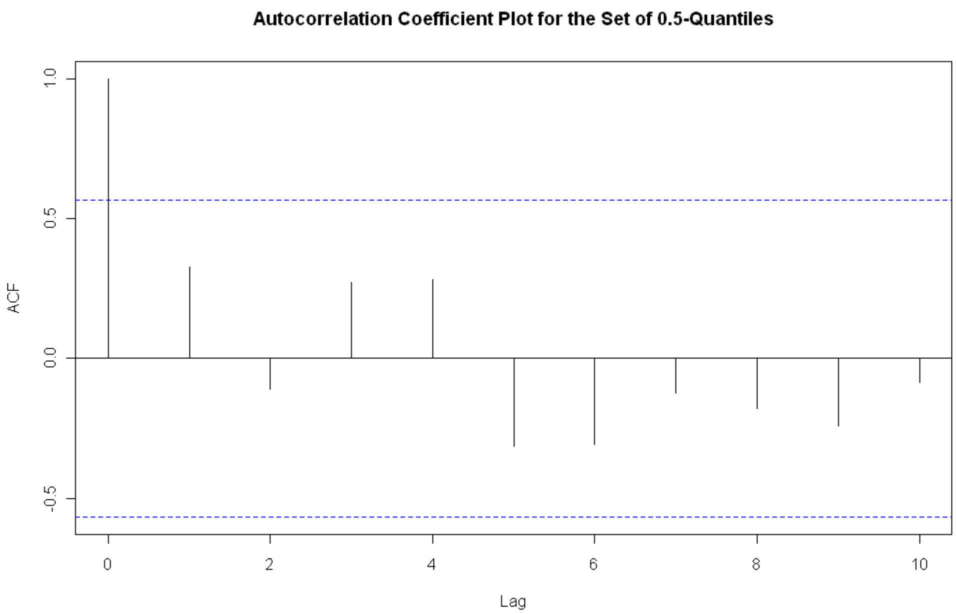


Figure 3.4: Autoregression Coefficients Graph for 0.5 Quantiles (Median) of GHP Times over Lags

to be approximately 0. The 95% confidence limits for the correlogram can be plotted at approximately $0 \pm 2/\sqrt{n}$; therefore, the approximate 95% confidence band in our case is $\pm 2/\sqrt{12} = \pm 0.5773503$. As one observes, the autocorrelation coeeficients we calculated are in this limit, hence, we conclude our observations are from a not autocorrelated population.
The above results let us conclude that there is no consistent pattern from one OP period to the other, i.e, the intervals 1-period apart are unrelated, as the autocorrelation coefficients are not significantly different from 0.

3.3 BASIC FORECASTING METHODS

3.3.1 Regression Analysis

Regression is a simple method which is used to analyze the reletionship between two variables, namely explanatory variable X and dependent variable Y. The goal of the method is to find the best fitting curve in order to predict Y from X. A linear regression line is of the form $Y = aX + b$, where a is the *slope* and b is the *intercept*. The aim of the algorithm is to adjust the values of the slope and the intercept in order to find the line which produces the best forecasts.
The most common method to fit the data into a curve is *Least Squares Estimates*. The result is reached by minimizing the sum of the squares of vertical distances of the points from the curve. If we think the observed data as a function of explanatory variable with an error, the following equation describes our model:

$$Y_t = f(X_t) + e_t \quad t = 1, 2, \ldots, n \tag{3.1}$$

where the linear function f and e_t determines the pattern and the error, respectively. The critical task on forecasting is to seperate the pattern from the error component so that the former can be used for forecasting.
If the function f in the equation (3.1) is quadratic, the method is called quadratic regression.

Equation of the Least-Squares Regression Line

The equation of the lest-squares regression line is given by

$$Y = aX + b$$

where

$$a = \frac{\sum_{i=1}^{n} X_i Y_i - \frac{\sum X_i \sum Y_i}{n}}{\sum_{i=1}^{n} x_i^2 - \frac{(\sum X_i)^2}{n}}$$

and

$$b = \bar{Y} - a\bar{X}$$

$\bar{X}$ and $\bar{Y}$ are the means of X and Y, respectively.

3.3.2 Exponential Smoothing

In this section, we describe a set of methods which assign weights to the observations. Since the weights exponentially decrease as the observations get older, these methods are called exponential smoothing procedures.
There are single, double and more complicated exponential smoothing methods.

Single Exponential Smoothing

The below equations give the general from of single exponential smoothing (SES):

$$\begin{aligned} F_1 &= Y_1 \\ F_{t+1} &= \alpha Y_t + (1-\alpha)F_t, \quad \text{where } \alpha \in (0,1), \quad t = 1, 2, \ldots, n \end{aligned} \tag{3.2}$$

F_t represents the forecast at time t while Y_t is the actual observation. Simply, one can use $\alpha = 1/n$ where n is the total number of available observations. However, there are more sophisticated methods to estimate α. The estimation involved in exponential smoothing is a non-linear optimization problem.
Single exponential smoothing algorithm does not need to store all of the historical data. It only requires to store the most recent observation, the most recent forecast and the value of α.

As shown by Muth [1960], single exponential smoothing predictor derived from the equation (3.2) is optimal if and only if Y_t is generated by the ARIMA$(0, 1, 1)$ (Auto-Regressive Integrated Moving Averages) process $(1-B)Y_t = [1-(1-\alpha)B]\epsilon_t$. On the other hand, if the data is stationary, one still obtains fairly good approximation, but when the existence of a trend, the SES method explained above is inadequate (Makridakis, Wheelwright, McGee[1983]).

Adaptive Response Rate Single Exponential Smoothing

Adaptive Response Rate Single Exponential Smoothing (ARRSES) is a single exponential smoothing method with dynamic model parameter. Smoothing parameter α_t depends on the step t in the following way

$$
\begin{aligned}
F_1 &= Y_1 \\
F_{t+1} &= \alpha_t Y_t + (1-\alpha_t)F_t, \quad \text{where } \alpha \in (0,1), \quad t = 1, 2, \ldots, n \\
e_t &= Y_t - F_t \\
E_t &= \beta e_t + (1-\beta)E_{t-1} \\
M_t &= \beta|e_t| + (1-\beta)M_{t-1} \\
\alpha_{t+1} &= \left|\frac{E_t}{M_t}\right|
\end{aligned} \tag{3.3}
$$

E_t is called smoothed error while M_t is absolute value of smoothed error.

Holt-Winters Three Parameter Exponential Smoothing

3.3.3 Box Jenkins Method

3.4 FORECAST ACCURACY

A fundamental concern while forecasting is how to measure the suitablity of a particular forecasting method. In most forecasting situations accuracy is seen like the criterion for selecting a forecasting method. In time-series modeling it is possible to use a subset of known data to forecast the rest of the known data which enables one to study the accuracy of the forecasts more directly.

There are many nave thecniques as well as mathematically sophisticated techniques. Some of forecast accuracy measures are stated below.

3.4.1 Standard Statistical Measures

If Y_i is the actual value in concern at time i and F_i is the forecasted (or fitted) value for the same time period then the error is defined:

$$e_i = Y_i - F_i$$

If there are observations and forecasts for n time periods then one can talk about the following statistical measures:

- Mean Error

  ME = $\sum_{i=1}^{n} e_i/n$

- Mean Absolute Error

  MAE =$\sum_{i=1}^{n} |e_i|/n$

- Sum of Squared Errors

  SSE = $\sum_{i=1}^{n} e_i^2$

- Mean Squared Error

  $MSE = \sum_{i=1}^{n} e_i^2/n$

- Standard Deviation of Errors

  SDE = $\sqrt{\sum_{i=1}^{n} e_i^2/(n-1)}$

A forecaster may want to see all of the measures above routinely, but the main points is to recognize the limitations of each. For instance, in most cases, the forecaster's aim is to minimize the mean squared error (or sum of square errors), however, this measure has two disadvantages. First of all, this measure refers to fitting a model to historical data. Such a fit does not need to give good forecasts for future. An MSE of value 0 can always be obtained by fitting a model to the data by using a function (or polynomial) of sufficiently high order or a suitable Fourier transformation. Overfitting a model to data is as bad as failing to identify the non-random pattern in the data.

Secondly, each different method has its own procedures in the fitting phase and this is related to the measure of accuracy. For example, regular linear regression method gives the same weight to the error while minimizing the MSE whereas Box-Jenkins method follows a non-linear optimization process. Therefore, comparing the accuracy of those methods on a single criterion, such as MSE, is of limited value.

3.4.2 Relative Measures

Because of the drawbacks of the above mentioned measures, alternative measures have been proposed. One can find those especially dealing with percentages below.

- Percentage Error

  $PE_i = \left(\frac{X_i - F_i}{F_i}\right)(100)$

- Mean Percentage Error

  $\text{MPE} = \sum_{i=1}^{n} PE_i/n$

- Mean Absolute Percentage Error

  $\text{MAPE} = \sum_{i=1}^{n} |PE_i|/n$

As one may imagine, among the above measures, MPE tends to be small as positive values and negative values would cancel each other. Therefore MAPE is introduced to get rid of this drawback. In many cases, knowing the mean absolute percentage error is more useful than knowing the mean squared error. In this paper, we use MAPE to measure the forecast accuracy and find the optimum model parameters.

3.5 NONPARAMETRIC CURVE ESTIMATION

In this section, in order to reach our aim, we followed a non-parametric robust approach. Denoting the OP periods by $t_k, k = 1, 2, \ldots, n$ (in our recent case $n = 12$) and τ sample quantiles for each OP period by Y_k^τ, one can construct the following non-parametric model:

$$Y_k^\tau = \theta^\tau(t_k) + \xi_k^\tau, \quad k = 1, 2, \ldots, n$$

where Y_k^τ is observed on a discrete grid of a closed bounded interval $I \subset \mathbb{R}$, ξ_k^τ are independent identically distributed random noises, $\theta(x) \in \Theta_\beta(H, L), x \in I$, with non-parametric class $\Theta_\beta(H, L)$ which will be defined below.

Without loss of generality, we assume the interval I to be unit, i.e, $I = [0, 1]$, as any closed bounded interval can be transformed into the unit interval.

In the above model, one should interprete Y_k^τ as observed τ sample quantiles while θ function gives the real τ quantiles at time t_k, which we cannot identify from the sample without knowing the exact distribution. ξ_k^τ is the error term whose τ quantile we assume to be 0.

The distribution of ξ_k^τ is absolutely continuous with density f_{ξ^τ}, which is assummed to be unknown. Moreover, we have the following assumptions on the density of ξ^τ with positive constans δ, p, L_{ξ^τ} and M_{ξ^τ}:

- (A1) $F_\xi(0) = \int_{-\infty}^{0} f_\xi(u)du = \tau$, i.e, τ quantile of the distribution of ξ_k^τ is 0, for all $0 < \tau < 1$.
- (A2) $\inf_{|x|<\delta} f_{\xi^\tau}(x) \geq p$, i.e, in the δ neighborhood of (any) x, the density of ξ^τ is strictly positive.
- (A3) $|f_{\xi^\tau}(0)| < \infty$

- (A4) The function $f_{\xi^\tau}(u)$ satisfies the Lipschitz condition:

$$|f_{\xi^\tau}(u_2) - f_{\xi^\tau}(u_2)| \le L_{\xi^\tau}|u_1 - u_2|, \qquad u_1, u_2 \in \mathbb{R}$$

In this study, the design $\{t_k\}$ is assumed to be almost equidistant. In other words, the following conditions are assumed on the design:

- (D1). $0 = t_0 \le t_1 \le \ldots \le t_n = 1$;
- (D2). $|t_l - t_m| \le D|l - m|/n$ for all $0 \le l, m \le n$.

The unknown function $\theta(x)$ on the interval $[0, 1]$ belongs to the Lipschitz function class $\Theta_\beta = \Theta_\beta(H, L)$ with smoothness β. That is, for some positive H, L and $0 < \beta \le 1$:

$$\Theta_\beta(H, L) = \{\theta : |\theta(0)| \le H, |\theta(u) - \theta(v)| \le L|u - v|^\beta, u, v \in [0, 1]\}.$$

Our aim is to estimate a function value for $\theta(x)$, for $x \in I$. An estimator $\hat{\theta}_n^\tau = \hat{\theta}_n^\tau(Y_0^\tau, Y_1^\tau, \ldots, Y_n^\tau)$ is an arbitrary measurable function of the observations.
We denote the frequency of the observations per unit interval by parameter n. We study the sequence of models:

$$Y_{k,n}^\tau = \theta^\tau(t_{k,n}) + \xi_{k,n}^\tau, \qquad k = 0, 1, \ldots, n.$$

For the sake of simplicity in the notation, we omit the subscript n, and superscript τ thus, the notation of our model is

$$Y_k = \theta(t_k) + \xi_k, \quad k = 0, 1, \ldots, n. \tag{3.4}$$

Observe that the conditional expectation does not necessarily exist under the assumptions (A1) - (A4), therefore, our model 3.3 is in general not a non-parametric regression. The aim of this study is predicting one step ahead function value of θ given historical observations and estimates of θ.

3.5.1 A Recursive Estimator

In this section, we try to find a recursive estimator based on a stochastic approximation procedure and derive its rate of convergence, as the frequency of the observations per unit interval, n, tends to infinity, based on the results achieved by Belitser and van de Geer [2000] ([5]).
Remeber the piecewise linear loss function:

$$\rho_\tau(u) = u\tau 1\{u \geq 0\} + u(\tau - 1)1\{u < 0\}$$

Now define the derivative of $\rho_\tau(u - v)$ with respect to v as follows:

$$d(u - v) = \frac{d\rho(u - v)}{dv} = -\tau 1\{u - v \geq 0\} - (\tau - 1)1\{u - v < 0\}$$

Moreover, for some fixed positive h, define the function:

$$S(u, v) = \begin{cases} -d(u - v) & \text{if } |v| \leq H + L + h \\ -v & \text{if } |v| > H + L + h \end{cases}$$

$$C_0 = 2/min\{p, p\delta/2(H + L) + h, 1/2\}$$

and the sequence $\gamma_n = n^{-2\beta/(2\beta+1)} \log n$, where the constants β, H and L appear in the definition of the class Θ_β.
For any $0 < \tau < 1$, the following recursive formula gives an estimator for the function value θ_k:

$$\hat{\theta}_{k+1} = \hat{\theta}_k + \gamma_n S(Y_k, \hat{\theta} k),\ k = 0, 1, \ldots, n - 1$$

with the initial value $\hat{\theta}_0 = \theta_1$.

In our case, the observations occur successively and at a time moment, t, the information based on only the observations occuring by t is available. In the above formula, since θ_k stands for $\theta(t_k)$, and t_k stands for the time-moment, this estimation procedure is a filtering algorithm and filtering algorithms are most appropriate in situations such as ours. Observe that, the estimator of the function θ_k is a measurable function with respect to $Y_1, Y_2, \ldots, Y_k$, i.e, the information available upto the time-moment t_k.

As one can see, we need to know a constant in the algorithm which must be bigger than $H + L$ (we took it as $H + L + h$). Although this seems to be restrictive feature of the algorithm, we can assume this without loss of generality, because we can prove the result with a sequence $H_n$converging to infinity sufficiently slowly instead of constant $H + L + h$, based on Belitser and de Geer[2000]([**?**, Bel]. This would make the proof of the result lengthier. In practise, one can use simply $S(u, v) = \tau 1\{u - v \geq 0\} + (1 - \tau)1\{u - v < 0\}$.

If we examine the above algorithm, we see that the expectation of $\hat{\theta}_{k+1}$ given the observations upto the time-monnment t_{k-1} is:

$$E(\hat{\theta}_{k+1}|Y_1, Y_2, \ldots, Y_{k-1}) = \hat{\theta_k} + \gamma_n(\tau - F_\xi(\hat{\theta}_k - \theta_k))1\{|\hat{\theta}_k| \leq H+L+h\} - \gamma_n \hat{\theta}_k 1\{|\hat{\theta}_k| > H+L+h\}$$

Because we assumed the τ quantile of the error term ξ is 0, this expectation shows the requirement for the algorithm to update the estimator $\hat{\theta_k}$ in the sense of shifting the estimator in the right direction, towards θ_k.

We would like to start with stating the following lemma:
Denote $g(u, \theta_k) = \mathbb{E}_\theta S(Y_k, u)$.

Lemma 3.5.1.1 *Assume that the conditions (A1) - (A4) are satisfied. For any* $u \in \mathbb{R}, n \in \mathbb{N}, 0 \leq k \leq n$, *the representation* $g(u, \theta_k) = -G(u, \theta_k)(u - \theta_k)$ *holds uniformly over* $\theta \in \Theta_\beta$ *for some function* $G(u, \theta_k)$, *such that* $G_1 \leq G(u, \theta_k) \leq G_2$ *with some positive constants* G_1 *and* G_2.

Proof. We try to prove the lemma in cases.
Case 1: $|u| \leq H + L + h$
Lets first compute the function $g(u, \theta_k)$. Remember, it is the expectation of the function $S(Y_k, u)$:

$$g(u, \theta_k) = \mathbb{E}_\theta S(Y_k, u) = \mathbb{E}_\theta - d(Y_k - u) = -\tau 1\{Y_k - u \geq 0\} + (\tau - 1)1\{Y_k - u = 0\}$$

$$= \mathbb{E}_\theta(\tau 1_{\{}Y_k - u \geq 0\}) + \mathbb{E}((\tau - 1)1\{Y_k - u < 0\})$$

Using the basic properties of expectation, one gets:

$$g(u, \theta_k) = \tau\mathbb{P}(Y_k - u \geq 0) + (\tau - 1)\mathbb{P}(Y_k - u < 0)$$

Now, say, $q = \mathbb{P}(Y_k - u \geq 0)$, which implies $1 - q = \mathbb{P}(Y_k - u < 0)$, and rewriting the above equation:

$$g(u, \theta_k) = \mathbb{E}_\theta S(Y_k, u) = \tau q + (\tau - 1)(1 - q) = \tau q - \tau q + 1 - q$$

$$= \tau - (1 - q) = \tau - \mathbb{P}(Y_k - u < 0)$$

Using the equation (1), $Y_k = \theta_k + \xi_k$, one gets:

$$\mathbb{E}_\theta S(Y_k, u) = \tau - \mathbb{P}(Y_k - u < 0) = \tau - \mathbb{P}(\xi_k < u - \theta_k) = \tau - F_\xi(u - \theta_k)$$

Now, it is left to prove that there exist a function $G(u, \theta_k)$ satisfying the conditions in the lemma with:

$$G(u, \theta_k) = \frac{-g(u, \theta_k)}{(u - \theta_k)} = \frac{F_\xi(u - \theta_k) - \tau}{(u - \theta_k)}$$

We know F_ξ, being a distribution function, is monotone increasing and by assumption (A1) $F_\xi = \tau$; therefore,
if $(u - \theta_k) < 0$, $F_\xi(u - \theta_k) \leq \tau$ and $F_\xi(u - \theta_k)/(u - \theta_k) \geq 0$, and
if $(u - \theta_k) > 0$, $F_\xi(u - \theta_k) \geq \tau$ and $F_\xi(u - \theta_k)/(u - \theta_k) \geq 0$.
Hence, $G(u, \theta_k)$ is always non-negative. Now, denote $(u - \theta_k)$ by v. If $|v| \leq \delta$, with constant δ from assumption (A2), we can use a Taylor series expansion to approximate F_ξ around 0 upto the first order term.

$$F_\xi(v) = F_\xi(0) + F'_\xi(v^*)(v - 0)$$

with $F'_\xi(v*) = f_\xi(v^*)$, where $|v^*| \leq |v| \leq \delta$. Using (A1), $F(0) = \tau$, we can express G as:

$$G(u, \theta_k) = \frac{\tau + f_\xi(v^*)v - \tau}{v} = f_\xi(v*)$$

Now we can conclude by using the assumption (A2) and (A3) that:

$$0 < p \le G(u, \theta_k) \le \sup_{|v| \le \delta} f_\xi(v) < \infty$$

If $|u - \theta_k| = |v| > \delta$, then

$$G(u, \theta_k) = \frac{F_\xi(u - \theta_k) - \tau}{u - \theta_k} \le \frac{1}{\delta}$$

Moreover, since $\theta \in \Theta$,

$$|\theta(u)| = |\theta(u)| - |\theta(0)| + |\theta(0)| \le |\theta(u) - \theta(0)| + |\theta(0)| \le L|u - 0|^\beta + H = Lu^\beta + H \le H + L$$

for any $u \in [0, 1]$. Using this result:

$$|u - \theta_k| \le |u| + |\theta_k| \le 2(H + L) + h$$

This leads us to:

$$G(u, \theta_k) = \frac{F_\xi(u - \theta_k) - \tau}{u - \theta_k} \ge \frac{min\{F_\xi(\delta) - \tau, \tau - F_\xi(-\delta)\}}{2(H + L) + h}$$

Lets again use Taylor expansions around 0 upto the first order tirm for $F_\xi(\delta)$ and $F_\xi(-\delta)$ to examine $min\{F_\xi(\delta) - \tau, \tau - F_\xi(-\delta)\}$.

$$F_\xi(\delta) = F_\xi(0) + F'_\xi(\delta^*)\delta \quad \text{with} \quad |\delta^*| \le |\delta|$$

$$\Rightarrow F_\xi(\delta) - \tau = \tau + f(\delta^*)\delta - \tau \ge p\delta \quad \text{by (A1) and (A2)}$$

and

$$F_\xi(-\delta) = F_\xi(0) + F'_\xi(\delta^{**})(-\delta) \quad \text{with} \quad |\delta^* * | \le |\delta|$$

$$\Rightarrow \tau - F_\xi(-\delta) = \tau - \tau + f(\delta^*)\delta \ge p\delta \quad \text{by (A1) and (A2)}$$

Therefore;

$$G(u, \theta_k) = \frac{F_\xi(u - \theta_k) - \tau}{u - \theta_k} \ge \frac{min\{F_\xi(\delta) - \tau, \tau - F_\xi(-\delta)\}}{2(H + L) + h} \ge \frac{p\delta}{2(H + L) + h} > 0$$

Case 2: $|u| > H + L + h$

In this case $S(Y_k, u) = -u$ so that $g(u, \theta_k) = \mathbb{E}_\theta S(Y_k, u) = -u$ which gives $G(u, \theta_k) = u/(u - \theta_k)$. Using the basic inequality $|u - \theta_k| \ge |u| - |\theta_k| \ge h$, conclude that:

$$G(u, \theta_k) = \frac{u}{u - \theta_k} = \frac{u - \theta_k + \theta_k}{u - \theta_k} = 1 + \frac{\theta_k}{u - \theta_k} \le |1| + \frac{|\theta_k|}{|u - \theta_k|} \le 1 + \frac{H + L}{h}$$

On the other hand, in order to find a lower bound for $G(u, \theta_k)$, lets look at the signs of u and θ_k:

- if u and θ_k are of the same sign then clearly $G(u, \theta_k) \geq 1$.
- if u and θ_k are of opposite signs, observe that,

$$|u - \theta_k| \geq H + L + h + |\theta_k| \ \ (\leq 2(H + L) + h)$$

then:

$$G(u, \theta_k) = \frac{u}{u - \theta_k} = \frac{u - \theta_k + \theta_k}{u - \theta_k} = 1 + \frac{\theta_k}{u - \theta_k} \geq 1 - \frac{|\theta_k|}{|u - \theta_k|} \geq 1 - \frac{|\theta_k|}{H + L + h + |\theta_k|}$$
$$= \frac{H + L + h}{H + L + h + |\theta_k|} \geq \frac{H + L + h}{2(H + L) + h} \geq \frac{1}{2}$$

To sum up;

- if $|u| \leq H + L + h \Rightarrow 1/2 \leq G(u, \theta_k) \leq 1 + [(H + L)/h]$
- if $|u| > H + L + h$
  - if $|u - \theta_k| \leq \delta \Rightarrow p \leq G(u, \theta_k) \leq \sup_{|x| \leq \delta} f_\xi(x)$
  - if $|u - \theta_k| > \delta \Rightarrow p\delta/2(H + L) + h \leq G(u, \theta_k) \leq 1/\delta$

Now we can take

$$G_1 = min\{p, p\delta/(2(H + L) + h), 1/2\}$$

and

$$G_2 = max\{1/\delta, \sup_{|x| \leq \delta} f_\xi(x), 1 + [(H + L)/h]\}$$

We proved the lemma.

■

Now we are ready to state and prove the theorem which specifies the rate of convergence of the estimator $\hat{\theta}_k$.

Theorem 3.5.1.2 *Suppose the conditions (A1) - (A4) are satisfied. Let* $K_n = \{k \in \mathbb{N} : C_0 n^{2\beta/(2\beta+1)} \leq k \leq n\}$*, where* C_0 *is defined by (algorithm). Then, for some* *positive constant C_1, any fixed α, and any sequence $\{k_n\}$ such that $k_n \in K_n$, the relations*

$$\sum_{n=1}^{\infty} P_\theta\{n^{\beta/(2\beta+1)}(\log n)^{(-3/2+\alpha)}|\hat{\theta}_{k_n} - \theta_{k_n}| > \epsilon\} < \infty$$

$$\limsup_{n\to\infty} \max_{k\in K_n} \frac{n^{2\beta/(2\beta+1)}}{(\log n)^2} \mathbb{E}_\theta(\hat{\theta}_k - \theta_k)^2 \leq C_1$$

hold uniformly over $\theta \in \Theta$

Proof. Introduce the following identities for convenience:
for any sequence $\{b_i\}$, $\sum_{i=m+1}^{m} b_i = 0$, and $\prod_{i=m+1}^{m} b_i = 1$.
Now define the differences $\delta_k = \delta_{k,n} = \hat{\theta}_k - \theta_k$ and $\Delta\theta_k = \theta_k - \theta_{k+1}$. Denote the difference $S(Y_k, \hat{\theta}_k) - g(\hat{\theta}_k, \theta_k)$ by $M(X_k, \hat{\theta}_k, \theta_k)$. We can, now, rewrite the (*algorithm*):

$$\hat{\theta}_k = \hat{\theta}_{k-1} + \gamma_n S(Y_{k-1}, \hat{\theta}_{k-1}) = (\theta_k - \theta_{k-1}) + \hat{\theta}_k - \theta_k = (\hat{\theta}_{k-1} - \theta_{k-1}) + \gamma_n S(Y_{k-1}, \hat{\theta}_{k-1})$$

$$\Rightarrow \delta_k = \delta_{k-1} + \gamma_n(M(Y_{k-1}, \hat{\theta}_{k-1}, \theta_{k-1}) + g(\hat{\theta}_{k-1}, \theta_{k-1})) + \Delta\theta_{k-1}$$

Using the relation, $g(\hat{\theta}_k, \theta_k) = G(\hat{\theta}_k, \theta_k)\delta_k$ from the LEMMA:

$$\delta_k = \delta_{k-1}(1 - \gamma_n G(\hat{\theta}_{k-1}, \theta_{k-1})) + \gamma_n M(Y_{k-1}, \hat{\theta}_{k-1}, \theta_{k-1}) + \Delta\theta_{k-1}$$

$k = 1, \ldots, n$. Observe that, for any k_0, $0 \leq k_0 \leq k$, iterating the above algorithm gives:

$$\delta_k = (\delta_{k-2}(1-\gamma_n G(\hat{\theta}_{k-2}, \theta_{k-2}))+\gamma_n M(Y_{k-2}, \hat{\theta}_{k-2}, \theta_{k-2}))(1-\gamma_n G(\hat{\theta}_{k-1}, \theta_{k-1}))+\gamma_n+\Delta\theta_{k-1}$$

$$= ((\delta_{k-3}(1 - \gamma_n G(\hat{\theta}_{k-3}, \theta_{k-3})) + \gamma_n M(Y_{k-3}, \hat{\theta}_{k-3}, \theta_{k-3}) + \Delta\theta_{k-3})(1 - \gamma_n G(\hat{\theta}_{k-2}, \theta_{k-2}))$$

$$+\gamma_n M(Y_{k-2}, \hat{\theta}_{k-2}, \theta_{k-2}) + \Delta\theta_{k-2})(1 - \gamma_n G(\hat{\theta}_{k-1}, \theta_{k-1})) + \gamma_n M(Y_{k-1}, \hat{\theta}_{k-1}, \theta_{k-1}) + \Delta\theta_{k-1}$$

.

.

.

$$= \delta_{k_0} \prod_{i=k_0}^{k-1} [1 - \gamma_n G(\hat{\theta}_i, \theta_i)] + \sum_{i=k_0}^{k-1} \{[\Delta\theta_i + \gamma_n M(Y_i, \hat{\theta}_i, \theta_i)] \prod_{j=i+1}^{k-1} [1 - \gamma_n G(\hat{\theta}_j, \theta_j)]\} \tag{3.5}$$

Now, remember *summation by parts* formula, which asserts:

$$\sum_{k=1}^{n} x_k z_k = X_n z_n - \sum_{k=1}^{n-1} X_k (z_{k+1} - z_k)$$

where $X_k = x_1 + x_2 + \ldots + x_k$.
Denoting $A_i = \sum_{j=k_0}^{i} a_j$, with $a_j = \Delta\theta_j + \gamma_n M(Y_j, \hat{\theta}_j, \theta_j)$, we can apply the summation by parts formula to the second part of the equation 3.4. Using the conventions we set, we get:

$$\sum_{i=k_0}^{k-1} \{[\Delta\theta_i + \gamma_n M(Y_i, \hat{\theta}_i, \theta_i)] \prod_{j=i+1}^{k-1} [1 - \gamma_n G(\hat{\theta}_j, \theta_j)]\} = \sum_{i=k_0}^{k-1} a_i \prod_{j=i+1}^{k-1} [1 - \gamma_n G(\hat{\theta}_j, \theta_j)]$$

$$= A_{k-1} - \sum_{i=k_0}^{k-2} A_i \gamma_n G(\hat{\theta}_{i+1}, \theta_{i+1}) \prod_{j=i+2}^{k-2} [1 - \gamma_n G(\hat{\theta}_j, \theta_j)] \tag{3.6}$$

Now, we can rewrite δ_k as:

$$\delta_k = \delta_{k_0} \prod_{i=k_0}^{k-1} [1 - \gamma_n G(\hat{\theta}_i, \theta_i)] + A_{k-1} - \sum_{i=k_0}^{k-2} A_i \gamma_n G(\hat{\theta}_{i+1}, \theta_{i+1}) \prod_{j=i+2}^{k-2} [1 - \gamma_n G(\hat{\theta}_j, \theta_j)]$$

The relation 3.5 holds for any sequence $\{a_i\}$, which allows us to take $a_{k_0} = 1$ and $a_j = 0$ for all $j > k_0$ and $A_i = \sum_{j=k_0}^{i} a_j = 1$ for all i, so:

$$\sum_{i=k_0}^{k-2} \gamma_n G(\hat{\theta}_{i+1}, \theta_{i+1}) \prod_{j=i+2}^{k-1} [1 - \gamma_n G(\hat{\theta}_j, \theta_j)] = 1 - \prod_{i=k_0+1}^{k-1} [1 - \gamma_n G(\hat{\theta}_i, \theta_i)] \leq 1$$

for sufficiently large n, by Lemma
As Belitser and de Geer pointed out,

$$|\hat{\theta}_k| \leq H + L + h + b, \quad b = \sup_{u \geq 1} u^{-2\beta/(2\beta+1)} \log u$$

Because $\hat{\theta}_k$ and θ_k are bounded as we showed, $|\delta_k|$ is bounded uniformly in k and $\theta \in \Theta_\beta$ as well.

Now we can write the relation for $|\delta_k|$ as follows:

$$|\delta_k| \le |\delta_{k_0}|(1-\gamma_n G_1^{k-k_0} + \max_{k_0\le i\le k}|A_i|$$

for sufficiently large n, uniformly in $\theta \in \Theta_\beta$, where G_1 is as defined in the proof of the LEMMA. The above inequality is equalivant to:

$$\delta_k^2 \le \delta_{k_0}^2(1-\gamma_n G_1)^{2(k-k_0)} + 2\delta_{k_0}(1-\gamma_n G_1)^{k-k_0}\max_{k_0\le i\le k}|A_i| + \max_{k_0\le i\le k} A_i^2$$

Using the definition of A_i:

$$\max_{k_0\le i\le k} A_i^2 = \max_{k_0\le i\le k}\{\sum_{j=k_0}^{i} a_j\}^2 = \max_{k_0\le i\le k}\{\sum_{j=k_0}^{i}\Delta\theta_j + \gamma_n M(Y_j,\hat{\theta}_j,\theta_j)\}^2$$

$$\le c\max_{k_0\le i\le k}\{\sum_{j=k_0}^{i}\Delta\theta_j\}^2 + C\max_{k_0\le i\le k}\gamma_n^2\left\{\sum_{j=k_0}^{i} M(Y_j,\hat{\theta}_j,\theta_j)\right\}^2$$

for constants c and C in $\mathbb{R}$. As in the research of Belitser and van de Geer [number], we choose the level of backward recursion k_0 as follows:

$$k_0 = k_0(k) = k_0(k,n) = \lfloor k - C_0\gamma_n^{-1}\log n\rfloor$$

Hence,

$$C_0\gamma_n^{-1}\log n \le k - k_0 \le C_0\gamma_n^{-1}\log n + 1$$

and therefore $k_0(k)$ are properly defined for those k which satisfy

$$k \in \{i \in \mathbb{N} : C_0\gamma_n^{-1}\log n \le i \le n\} = J_n$$

Now, let us derive a bound for $(1-\gamma_n G_1)^{k-k_0}$:

$$(1-\gamma_n G_1)^{k-k_0} \le C(1-\gamma_n G_1)^{C_0\gamma_n^{-1}\log n} \le Ce^{-2\log n} = Cn^{-2}$$

for sufficiently large n as long as C_0 satisfies $C_0 G_1 \ge 2$.
Moreover:

$$\max_{k_0\le i\le k}\{\sum_{j=k_0}^{i}\Delta\theta_j\}^2 \le \max_{k_0\le i\le k} L^2(x_i - x_{k_0})^{2\beta} \le c(k-k_0)^{2\beta}/n^{2\beta}$$

uniformly in $\theta \in \Theta$. In order to evaluate $C\gamma_n^2\{\sum_{j=k_0}^{i} M(Y_j, \hat{\theta}_j, \theta_j)\}^2$, observe that $\{M(Y_j, \hat{\theta}_j, \theta_j)\}_{j=k_0}^{k}$ is a *martingale difference* with respect to the natural filtration $\{\mathcal{A}_j\}_{j=k_0}^{k}$, where $\mathcal{A}_j = \sigma(Y_0, Y_1, \ldots, Y_j)$, the σ-algebra generated by $Y_0, Y_1, \ldots, Y_j$. Lets first remind the definition of martingale:

Definition: A stochastic process $M = (M_n)_{n\geq 0}$ is called a *martingale* with respect to the filtration $\mathcal{F}$ if it satisfies the following conditions:

- (1)M is adapted to $\mathcal{F}$
- (2)$\mathbb{E}(|M_n|) < \infty$ for all $n \geq 0$
- (3)$\mathbb{E}(M_{n+1}|\mathcal{F}_n) = M_n$ a.s. $\forall n$

First, observe from the definition of M and the bound of $|\hat{\theta}_k|$ that $|M(Y_j, \hat{\theta}_j, \theta_j)| \leq 2\max\{1, H + L + h + b\}$, almost surely. Moreover,

$$\begin{aligned}\mathbb{E}(M(Y_{j+1}, \hat{\theta}_{j+1}, \theta_{j+1})|\mathcal{A}_j) &= \mathbb{E}(S(Y_{j+1}, \hat{\theta}_{j+1})|\mathcal{A}_j) - \mathbb{E}(g(\hat{\theta}_{j+1}, \theta_{j+1})|\mathcal{A}_j) \\ &= 0, \quad j = 1, 2, \ldots, n-1. \end{aligned} \tag{3.7}$$

Therefore, $\{\sum_{j=k_0}^{m} M(Y_j, \hat{\theta}_j, \theta_j)\}_{m=k_0}^{k}$ is a martingale. Now, remember Doob's Submartingale Inequality:

Doob's Submartingale Inequality: Let Z be a martingale or non-negative submartingale. Then, for $c > 0$ and $p > 1$,

$$\mathbb{P}(\max_{k\leq n} Z_k \geq c) \leq c^{-1}\mathbb{E}(Z_n) \tag{3.8}$$

and

$$\|\max_{k\leq n} Z_k\|_p \leq \frac{p}{p-1}\|Z_n\|_p \tag{3.9}$$

The special case where $p = 2$ gives

$$\mathbb{E}[(\max_{k\leq n} Z_k)^2] \leq 4\mathbb{E}[Z_n^2]$$

Applying Doob's Submartingale Inequality to $\sum_{j=k_0}^{k} M(Y_j, \hat{\theta}_j, \theta_j)$, we obtain:

$$\mathbb{E}\{\max_{k_0 \le i \le k} [\sum_{j=k_0}^{i} M(Y_j, \hat{\theta}_j, \theta_j)]^2\} \le 4\mathbb{E}\{\sum_{j=k_0}^{k} M(Y_j, \hat{\theta}_j, \theta_j)\}^2 = C(k - k_0)$$

uniformly in $\theta \in \Theta_\beta$ and over $k \in J$ for some constant C.

From this point on, following the same steps as Belitser and de Geer, one concludes:

$$\begin{aligned} \phi_n[(1-\gamma_n G_1)^{k-k_0} + \max_{k_0 \le i \le k} |\sum_{j=k_0}^{i} \Delta\theta_j|] &\le \frac{C\phi_n}{n^2} + \frac{L\phi_n(k-k_0)^\beta}{n^\beta} \\ &\le \frac{c}{(\log n)^{3/2+\alpha}} \end{aligned} \quad (3.10)$$

uniformly in $\theta \in \Theta_\beta$ and over $k \in J_n$, where $\phi_n = n^{\beta/(2\beta+1)}(\log n)^{-3/2+\alpha}$.

Furthermore, using the Azuma-Hoeffding inequality, one has the following bound:

$$\begin{aligned} \mathbb{P}\{c\phi_n\gamma_n \max_{k_0 \le i \le k} |\sum_{j=k_0}^{i} M(Y_j, \hat{\theta}_j, \theta_j)| > \epsilon/2\} &\le 2\exp\{-\frac{Cn^{2\beta/(2\beta+1)}(\log n)^{1+2\alpha}}{k-k_0}\} \\ &\le 2\exp\{-c(\log n)^{1+2\alpha}\} \le Cn^{-2} \end{aligned} \quad (3.11)$$

uniformly in $\theta \in \Theta_\beta$ and over $k \in K_n$. Uniformness of the above results lead us to

$$\sum_{n=1}^{\infty} \mathbb{P}\{\phi_n|\delta_{k_n}| > \epsilon\} < \infty$$

uniformly in $\theta \in \Theta_\beta$ for any sequence $\{k_n\}$ such that $k_n \in K_n$. The result then follows. ■

Remark

As Belitser and de Geer[2000]([5]) state, the results hold if we allow ξ_k's to have different distributions, all satisfying the conditions (A1) - (A4). Furthermore, defining the class $\mathcal{P}_\xi = \mathcal{P}_\xi(\delta, p, M)$ of densities satisfying (A1) - (A4) with the condition $\sup_{u\in\mathbb{R}} |f(u)| \le M$, one can show that all the results hold uniformly overall joint distributions of $(\xi_0, \xi_1, \ldots, \xi_n)$, for all independent ξ with densities $f_\xi \in \mathcal{P}_\xi$.

3.5.2 Estimation of γ_n

Estimation of γ_n is a nonlinear optimization problem. One has to choose an accuracy measure to minimize in order to decide on γ_n. In this study we chose this measure as the mean square error.
Experiments showed the estimations using a parameter depending on the number of observations, n, does not lead to accurate results. (The optimization, estimation results and the conclusion will be written later.)
We try to improve our model parameters.

3.5.3 Improvement of the Model Paramters

In this section, we try to find a parameter, namely γ_k, depending on the iteration step for the model

$$\hat{\theta}_{k+1} = \hat{\theta}_k + \gamma_k S(Y_k, \hat{\theta}k)$$

with the initial value $\hat{\theta}_0 = \theta_1$, for any $0 < \tau < 1$.
We combine the method of ARRSES in section 3.3.2. There is a significant improvement in the estimations. (Results will be included later).

3.6 EVALUATION OF FORECAST RESULTS

3.7 COMPARISON AND COMBINATION OF FORECAST METHODS

Bibliography

[1] van der Vaart A.W., *Asymptotic Statistics*, Cambridge Series in Statistical and Probabilistic Mathematics, (1998).

[2] Williams D., *Pobability with Martingales*, Cambridge Mathematical Textbooks, (1991).

[3] Bowman K.O., Shenton L.R., *Properties of Estimators for the Gamma Distribution*, Union Carbide Corporation Nuclear Divison, Oak Ridge, Tennessee (1968).

[4] Geyer C.J., *Maximum Likelihood in R*, (2003).

[5] Belitser E., van de Geer S., *On Robust Recursive Nonparametric Curve Estimation*, High Dimensional Probability II(2000), pp. 391-404.

[6] Rice J.A., *Mathematical Statistics and Data Analysis*, Wadsworth and Brooks, Pacific Grove, California (1988).

[7] Makridakis S., Wheelright S., McGee V., *Forecasting: Methods and Applications*, John Wiley and Sons, Inc., 2nd edition, (1983).

[8] Granger C.W.J., Newbold P., *Forecasting Economic Time Series*, Academic Press, Inc., (1977).

[9] Koenker R., *Quantile Regression*, Cambridge University Press, New York, (2005).

[10] Balakrishnan N., Kateri M., *On the Maximum Likelihood Estimation of Parameters of Weibull Distribution Based on Complete and Censored Data*, Elsevier Statistics and Probability Letters 78 (2008), pp. 2971-2975.

[11] van den Boom L., *STIPT B737 Simulation Tool Insight in Performance Turnaround Boeing 737*, Internship Report, KLM Royal Dutch Airlines, Ground Services, Capacity Planning and Management, (2003).

Appendix A

BACKGROUND

A.1 GAMMA DISTRIBUTION

The *Gamma Distribution* is a two-parameter continuous distribution which is widely used in statistical applications, such as life testing procedures, process time examinations and whether modification experiments. It has a scale parameter λ (which is equivalent to a rate parameter $\frac{1}{\lambda}$) and a shape parameter k.
The probabilty density function of Gamma distribution with shape k and scale λ is given by:

$$f(x) = x^{k-1}\frac{e^{(-x/\lambda)}}{\Gamma(k)}\lambda^{-k}, \text{ for } 0 \leq x < \infty \text{ and } k, \lambda > 0.$$

where $\Gamma(z) = \int_0^\infty t^{z-1}e^{-t}dt$.
If k is an integer, the distribution represents the sum of k independent exponentially distributed random variables with mean λ.
The cumulative distriubtion function of Gamma distribution with parameters k and λ is:

$$F(x) = \int_0^x f(u)du = \frac{\gamma(k, x/\lambda)}{\Gamma(k)}$$

where γ is the incomplete gamma function, $\gamma(s,t) = \int_0^t u^{s-1}e^{-u}du$.
The mean and the variance of a Gamma(k, λ) distributed random variable are

$$\mu = k\lambda \quad \text{and} \quad \sigma^2 = k\lambda^2,$$

respectively.

A.2 WEIBULL DISTRIBUTION (2-parameter)

Like the Gamma distribution, 2-parameter Weibull distribution is a continuous probability distribution. It has a shape parameter k and a scale parameter λ. It is widely used to analyse life data. The shape parameter k is related with the failure rate as follows:
If the failure rate decreases by time then $k < 1$, if it increases, $k > 1$. If the failure rate is constant over time $k = 1$.
The probability density function of a Weibull random variable with shape k and scale λ is given by

$$f(x) = \frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1} e^{-(x/\lambda)^k}, \quad x \geq 0, \ \ k > 0, \ \ \lambda > 0$$

The cumulative distribution function is

$$F(x) = 1 - e^{-(x/\lambda)^k}$$

for $x \geq 0$ with $k > 0$ shape and $\lambda > 0$ scale parameters.
The mean of a random variable, X, distributed according to Weibull distribution with shape k and scale λ is:

$$\mu = \lambda\Gamma\left(1 + \frac{1}{k}\right)$$

where $\Gamma(z)$ is the previously defined gamma function. Similarly, the variance of X is given by

$$\sigma^2 = \lambda^2\Gamma\left(1 + \frac{2}{k}\right) - \mu^2$$

A.3 MAXIMUM LIKELIHOOD ESTIMATION

Maximum Likelihood Estimation (MLE) is a statistical method which is used to fit a statistical model to the given data providing model's parameters.

Suppose that random variables $X_1, X_2, \ldots, X_n$ have a joint density or frequency function $f(x_1, x_2, \ldots, x_n|\theta)$. Given observed values $X_i = x_i$, $i = 1, 2, \ldots, n$, the likelihood of θ as a function of $x_1, x_2, \ldots, x_n$is defined as

$$L(\theta) = f(x_1, x_2, \ldots, x_n|\theta),$$

$\theta \in \Theta$, where Θ is the parameter space. Observe that the likelihood is a function of θ rather than x_i. If the distribution is discrete, which means f is a frquency function, the likelihood function gives the probability of observing the given data as a function of the parameter θ. The maximum likelihood estimate of θ is that value of θ maximizing the likelihood, i.e making the observed data "most probable" or "most likely" (Rice [year]).

The likelihood function for n independent, identically distributed random variables $(X_1, \ldots, X_n)$ is given by

$$L(\theta) = \prod_{i=1}^{n} f(X_i|\theta)$$

where $\theta =\in \Theta$, where f is the marginal density function of X_i.

The above likelihood function does not have to have a maximizer. Even if it does have a maximizier, that need not to be unique. In the case that this function has more than one maximizer, most people tend to take the global maximizer as estimated parameters, however in general a local maximizer may give better results than the global maximizer. Therefore, an optimization algorithm which is initialized at a good starting point and aims to find a local maximizer would yield MLE parameters, if it converges to a solution. A good starting point, theoretically, obeys the square root law, that is the estimation error goes to 0 with a rate $c/n^{1/2}$, where c is a constant. One may use the sample parameters as a starting point of the algorithm, as those are the best one can get from the sample. (Maximum Likelihood in R, Charles J Geyer [2003])

In order to get rid of the divergence problems one may take suitable transformations of likelihood function. Because the parameter θ which maximizes L is equal to the parameter which maximizes log(L), as maxima are not affected by monotone transformations, one calculates log(L), that is the log-likelihood function:

$$l(\theta) = \sum_{i=1}^{n} \log\left[f(X_i|\theta)\right]$$

A.3.1 Maximum Likelihood Estimator for Gamma Distribution

If X follows a Gamma distribution with $\theta = (k, \lambda)$ where k is the shape and λ is the scale parameters, then

$$l(\theta) = (k-1)\sum_{i=1}^{n} \log(x_i) - \sum_{i=1}^{n} \frac{x_i}{\lambda} - nk\log(\lambda) - n\log(\Gamma(k)) \tag{A.1}$$

given $X_i = x_i$.

Now it is straightforward to maximize the log-likelihood function $A.1$ with respect to k and λ by taking the derivative of $l(\theta)$ and equaling it to 0.

First, taking the derivative with respest to λ gives:

$$\frac{dl(k,\lambda)}{d\lambda} = \frac{\sum_{i=1}^{n} x_i}{\lambda^2} - \frac{nk}{\lambda}$$

Equaling the above equation and solving for λ yields the maximum likelihood estimator of this parameter:

$$\hat{\lambda} = \frac{1}{kn}\left(\sum_{i=1}^{n} x_i\right)$$

If one substitutes the estimator of λ in the equation $A.1$, it gives:

$$l_{\hat{\lambda}}(k) = (k-1)\sum_{i=1}^{n} l\log(x_i) - nk - nk\log\left(\sum_{i=1}^{n} x_i/kn\right) - n\log(\Gamma(k))$$

Taking the derivative with respect to k and setting it equal to 0 yields the nonlinear equation

$$\log(\hat{k}) - \psi(\hat{k}) = \log\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) - \frac{1}{n}\left(\sum_{i=1}^{n} \log x_i\right)$$

where $\psi(\hat{k}) = \Gamma'(\hat{k})/\Gamma(\hat{k})$ is the digamma function. There is no straightgorward calculation for $\hat{k}$, the estimator for the shape k, but numerical approaches are used to obtain an estimator.

The maximum likelihood estimators $(\hat{k}, \hat{\lambda})$ are known to be sufficient estimators for Gamma distribution (Bowman and Shenton [1968]).

A.3.2 Maximum Likelihood Estimator for Weibull Distribution

If X follows a Weibull distribution with $\theta = (k, \lambda)$ where k is the shape and λ is the scale parameters, then

$$l(\theta) = n \log k + (k-1) \sum_{i=1}^{n} \log x_i - kn \log \lambda - \sum_{i=1}^{n} \left(\frac{x_i}{\lambda}\right)^k \tag{A.2}$$

given $X_i = x_i$.
Taking the derivative of log-likelihood function with respect to λ gives:

$$\frac{dl(k,\lambda)}{d\lambda} = -\frac{kn}{\lambda} + \sum_{i=1}^{n} k \left(\frac{x_i}{\lambda}\right)^{k-1} \frac{x_i}{\lambda^2} \tag{A.3}$$

Equaling $A.3$ to 0, one obtains

$$\begin{aligned} -kn + \frac{1}{\lambda^k} k \sum_{i=1}^{n} x_i^k &= 0 \\ n\lambda^k = \sum_{i=1}^{n} x_i^k \Rightarrow \lambda^{k-1} &= \frac{\sum_{i=1}^{n} x_i^k}{n} \\ \Rightarrow \hat{\lambda} &= \left(\frac{\sum_{i=1}^{n} x_i^k}{n}\right)^{1/k} \end{aligned}$$

Taking the derivative of $l(\theta)$ with respect to k and equaling it to 0 with $\lambda = \hat{\lambda}$ yields

$$\frac{n}{k} + \sum_{i=1}^{n} \log x_i - \frac{n}{\sum_{i=1}^{n} x_i^k} \sum_{i=1}^{n} x_i^k \log x_i = 0 \tag{A.4}$$

We can now rewrite $A.4$ as follows

$$\frac{1}{k} = \frac{\sum_{i=1}^{n} x_i^k \log x_i}{\sum_{i=1}^{n} x_i^k} - \frac{1}{n} \sum_{i=1}^{n} \log x_i \tag{A.5}$$

Balakrishnan and Kateri[10] showed that the plots of RHS and LHS of $A.5$ would intersect exactly once, at the MLE of shape k. Therefore, $A.5$ leads us the unique maximum likelihood estimator of shape k.
Similar with Gamma distribution, the estimator for shape $\hat{k}$ for Weibull distribution is calculated by numerical methods.

A.3.3 Asymptotical Convergence of MLE

(*A.W. van der Vaart, Asymptotic Statistics (Cambridge Series in Statistical and Probabilistic Mathematics) (1998)*) In practise, in many cases, a set of independent identically distributed random variables is used to estimate the distribution parameters. In such cases, one is interested in the quality of the estimator, that is the convergence of the estimated parameters to the true parameters as the number of observations increases to infinity. In literature, this is reffered as asymptotic behaviour of the estimator. In a certain sense, Maximum Likelihood Estimator is asymptotically optimal. The behaviour of MLE is listed below:

The MLE is asymptotically unbiased, that is, its bias tends to 0 as the sample size increases to infinity, where bias of the estimator $\hat{\theta}$ is: $\mathbb{E}(\hat{\theta}) - \theta$.

The MLE is asymptotically efficient, that is, it achieves the *Cramer-Rao lower bound* when the sample size tends to infinity. This means that no asymptotically unbiased estimator can give better estimator than MLE in the sense of asymptotic mean squared error.

The estimator is asymptotically normal, that is, as the sample size increases, the distribution of the MLE tends to Gaussian with mean θ and the inverse of the *Fisher Information matrix* as covariance matrix.

We have to assume some regularity conditions to have the listed behaviours of the MLE. Those conditions are:

- The first and second derivatives of the log-likelihood function must be defined.
- The Fisher information matrix must not be zero, and must be continuous as a function of the parameter.
- The maximum likelihood estimator is consistent, i.e, as the sample size goes to infinity, the estimator converges to the true value of the parameter *in probability*.

Although the theory does not determine how large sample is required to have good degree of approximation, in practise the MLE appears to be approximately true if the sample size is moderately large.

A.4 CHI-SQUARE GOODNESS OF FIT TEST

Chi-Square goodness of fit test is used to test if the data comes from a particular distribution. Therefore, the null hypothesis is
$H_0 : F = F_0$
and the alternative hypothesis is
$H_1 : F \neq F_0$
where F is the sample distribution and F_0 is the particular distribution to be tested if the sample data comes from.
The test requires that the data first be grouped. The actual number of observations in each group is compared to the expected number of observations according to F_0 and the test statistics is calculated as a function of this difference, namely

$$\chi^2 = \frac{\sum_{i=1}^{k} (O_i - E_i)^2}{E_i} \tag{A.6}$$

where k is the number of bins the data is divided into, O_i is the actual number of observations in each bin i, and E_i is the expected number of observations in bin i with respect to the distribution F_0.
$A.6$ approximately follows the χ^2 distribution with $k - 1 - l$ degrees of freedom where l is the number of distribution parameters estimated.
The expected frequency is calculated by
$E_i = n(F(Y_u^i) - F(Y_l^i))$ where F is the cumulative distribution function for the distribution being tested, Y_u^i is the upper limit for the class i, Y_l^i is the lower limit of the class i, and n is the sample size.
The Chi-Square goodness of fit test is sensitive to the choice of bins. The test power will be affected by the number of groups, how they are defined, and the sample size. There is no optimal choice for the bin width since the optimal bin width depends on the distribution.

Appendix B

APPLICATIONS WITH DIFFERENT TYPES OF KLM GROUND HANDLING PROCESS TIMES DATA