Bachelor thesis
Computer Science

Radboud University

A Linear Quantile Mixed Regression Model for Prediction of Airline Ticket Prices

Author:
Tim Janssen
S4150880

First supervisor/assessor:
TMH Dijkstra
t.dijkstra@science.ru.nl

Second supervisor:
Saiden Abbas
saiden.abbas@cs.ru.nl

Second assessor:
Prof. Dr. Allard C.R. van Riel
a.vanriel@fm.ru.nl

August 3, 2014

Abstract

We find it frustrating that different passenger on the same flight in the same flight class pay very different prices for their tickets while getting the exact same service. This research proposes four statistical regression models for airline ticket prices and compare the goodness of fit. With this prediction model passengers can make a more informed decision whether to buy the ticket or wait a little longer. We used a data set containing 126,412 observations of ticket prices of 2,271 different flights from San Francisco Airport to John F. Kennedy Airport, these observations have been made on a daily basis by Infare [2]. We find a model that fits the behavior of the data fairly well many days before departure. Therefore this approach could help future air travelers to decide whether to buy a ticket or not.

Contents

1 Introduction **2**
1.1 Motivation . . . 2
1.2 Background . . . 3

2 Method **5**
2.1 Data description . . . 6
2.2 Linear Quantile Mixed Model . . . 8

3 Results **12**

4 Discussion **16**

5 Conclusions **17**
5.1 Future Work . . . 17
5.2 Acknowledgements . . . 18

A Appendix **20**

Chapter 1

Introduction

Corporations with a ”standing inventory” often use complex and dynamic policies to determine optimal prices for their products in order to maximize revenue [11]. Airlines are one branch of these companies, having the available seats on a plane as their standing inventory. They divide these seats into several buckets, where each bucket has its own fare price. Airlines rearrange these seats across the buckets to make more money out of them, this creates changes in the prices which customers have to pay for the flight. Such that different customers pay different prices for tickets of the same flight.

1.1 Motivation

Imagine yourself flying from airport **A** to airport **B** , you’re sitting on a seat next to the aisle while having other travelers next to you near the window. While on the flight you start a conversation with your neighbor about the flight you’re on and the prices you paid for this trip, only to find out that your conversation partner has paid $50,- less than you did for exactly the same services, what a bummer.
We find it frustrating that consumers not only pay such a difference in price while getting the exact same service, but also that they perhaps both could have saved some money if they had some knowledge about the behavior of the prices of these airline tickets. What’s even more frustrating is that as a consumer there is just so little you can do to fill the knowledge gap, without spending a significant amount of time in checking if the prices increase or decrease. Even when customers compare and buy tickets for the cheapest price trough the available comparing tools on the World Wide Web, it is still possible that on that specific day prices are more expensive than usual without you even noticing.

1.2 Background

Since prices of airline tickets change over time it can be a very lucrative business to predict when ticket prices are cheap. There have been several papers about the topic of predicting ticket prices and/or buy-wait strategies. In this section we will discuss the source of the problem (changing prices) and the possible methods that are currently available to handle this problem. Airlines manage their standing inventory with the use of yield management, by changing prices up or down they try to increase their yield based on for example historical demand and airplane capacity. Traditionally this is done by hand, now largely taken over by Yield Management Systems (YMS). An YMS basically tries to sell the right seat to the right customer for the right price at the right time such as they can maximize yield, or revenue. Although the airline industry is considered the birthplace of yield management it is not only applicable to airlines [10], other types of fields include but are not limited to hotel bookings, ship cruises and car rental. Of course specific adjustments have to be made to fit these systems to the specific field. For example in the airline industry: when traveling from airport **A** to airport **B** without a direct route available, you will have to travel trough a hub called airport **C** but another customer would like to go from **A** to **B** . An YMS will try to offer you both a competitive price while still maximize yield taken into account the throughput of such a hub-and-spoke network [4].
As an YMS tries to predict demand by using historical data and adjust prices based on it, research has been done in trying to predict ticket prices by using historical data, in [5] they applied a multi-strategy data mining technique called HAMLET on web-crawled airfare data. This research showed that it is possible to save costs for consumers by using data mining to crawled data from the internet where key variables, such as the number of seats, are missing. It uses time series analysis, reinforcement learning and as well rule based learning and produces a wait/buy advice as its output. Based on this paper, [12] has suggested another approach, regarding the theory of (marked) point process [9] and the random tree forest algorithm [3], which should have less computational difficulties than the HAMLETT approach. Its results show that they perform almost as well as HAMLETT does but does have a more useful prediction due to a given confidence and an possible interpretation of the prediction.
While these papers use data mining classification techniques to predict airline prices or to produce a buy/wait advice, there has also been research about predicting ticket prices using a statistical regression model, more specifically in [8] they propose a lag scheme model which shows that there are possibilities to reduce costs for customers given sufficient publicly-observable information. In this paper we suggest a novel approach for predicting airline prices using linear quantile mixed models. The idea behind

this approach is that we are not interested in how the average ticket prices behaves but are only interested in the lowest ticket prices, call them the real bargains. These bargains are much more interesting for customers than the average price behavior as they will deliver the highest cost reduction. We give four possible models for the regression and compare the goodness of fit.

Chapter 2

Method

Airline companies indicate their flights with a combination of letters and numbers such as "DL1940", which represents flight 1940 flying for Delta Airlines (DL), or "AS531" which is flight 531 flying for Alaska Airlines (AS) these are called flight numbers and usually indicate flights flying on a specific route on a specific time of the day. While this specific flight only flies once, other flights with the same flight number fly almost daily throughout the year. Tickets for these flight are available many days before the flight departure date, such that consumers can buy tickets in advance. Luckily for us this gives us also the possibility to observe the prices of each of these flights throughout, for example, 60 days before the flight leaves. An example plot of the price of these observations on the 60 days before departure is given below in figure 2.1.

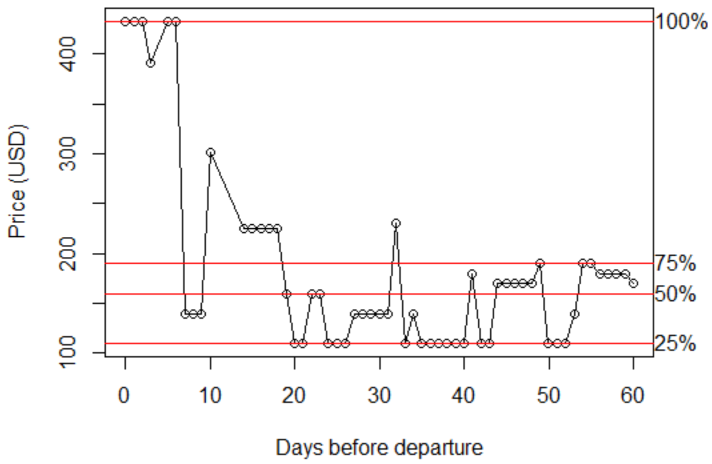

Figure 2.1: Observations of ticket prices of flight DL1940 departing from SFO heading to JFK at 2012-03-27. With (in red) four quantiles: $25^{th}, 50^{th}, 75^{th}$ and 100^{th}.

As we can see in this plot the price changes several times during the 60 before departure. If we look at figure 2.1 we think that the amount of days till departure is an important factor in determining the prices because of the YMS trying to fill an airplane many days before departure and reacting on demand/supply during the period before departure. Other factors of importance could be historical demand of previous flights, the day of the week, holidays, seats available in the plane et cetera. Unfortunately airline company's do not share all of this information to the public, but rather keep it for themselves. In figure 2.1 there are also four lines drawn, each line indicates a quantile. From bottom to top respectively the $25^{th}, 50^{th}, 75^{th}$ and 100^{th} quantile. All price observations which are below the 25^{th} quantile line belong to the cheapest 25% of prices. These prices are the one's we want to buy as they are cheap.

2.1 Data description

The data used in this research is collected by Infare Solutions [2], a Danish company who collects pricing airline data. This data has been bought and shared with us by Flyr [1] a company who specializes in the prediction of prices of airline tickets in the United States. The first observation in this dataset has been made on 2012-01-01 and from then on captures daily observations of each flight till 2013-05-31 . For the purpose of this research we will restrict ourselves to economy class tickets on single trip non-stop flights from San Francisco Airport (SFO) to John F. Kennedy airport (JFK). This enables us to specifically look into the behavior of price fluctuation without adding complexity of the influence of stops and business class demand.
Furthermore in this research we will be only looking at observations which are within 60 days of departure. Since the observations made in this data set are on a daily basis, we can have a maximum of 61 observations for every departing flight (these are 60 days observations days and the day of departure). Unfortunately some observations are missing in this data set, to have sufficient observations for each flight we chose to only include flights which have more than 30 observations. The remaining data that fit these prerequisites, consists of 126,412 observations of a total of 2,271 unique departing flights from 6 different Airlines. In table 2.1 we show the flight numbers which we will use throughout this paper, the number of flights we observed which had this flight number, the total number of observations we have of this flight and in the last column the percentage of missing observations. As we observed 61 days we can have a maximum of $61 * No.flights$ observations, say maxObs. The percentage is then calculated as: $(1 - \frac{No.Observations}{maxObs}) * 100$

Airline	*Flight Number*	*Departing Flights*	*No. Observations*	*Missing obs.*
American Airlines	AA12	165	9,298	7.5%
United Airlines	UA286	461	26,417	6.1%
Delta Airlines	DL1940	314	18,121	5.4%
JetBlue	B6648	498	28,843	5.1%
Virgin America	VX22	517	30,830	2.2%
US airways	US6634	316	12,903	33.1%
Sum		2,271	126,412	8.7%

Table 2.1: For every of the 6 airlines a specific flight number, the number of departing flights, the number of observations for this flight number and the percentage of missing observations.

For each of the observations made 27 different variables are stored. In our research we are only interested in 4 variables: departure date, observation date, price (including tax) and flight number. From these variables we derived 2 new variables, namely: is the flight departing in the weekend or is it a weekday and the number of days left to departure. An example of such a observation is shown in table 2.2.

Variable	*Value*
Flight number	DL1940
Departure Date	2012-05-05
Observation Date	2012-03-21
Price inc.	$ 101.40
Days left	45
Day of Week	Saturday
Is weekday	weekend

Table 2.2: A observation of price (incl), flight no, departure date of a specific flight, derived variables: days left till departure, day of week and is it weekend.

While looking into our data we've found several patterns that seems to apply to every flight. These are 1) There is a big price difference between the minimum and the maximum price you can pay for a ticket 2) Price usually tend to go up when nearing the day of departure 3) The day of the departure matters when looking at price.

1) When looking at the price behavior of a flight 60 days in advance as

for example in figure 2.1 we find that prices fluctuate strongly till the period before flight departure. Table 2.3 also shows two different flights having different flight numbers and the difference between the extreme prices (lowest versus highest).

Flight Number	*Departure Date*	*Min. Price*	*Max. Price*	*Difference*
DL1940	2012-05-05	$ 119.8	$ 504.8	$ 385
AA16	2012-12-24	$ 188	$ 661	$ 473

Table 2.3: Table 2.3: Minimum price, maximum price,for two different flights on the same route.

It are these fluctuations in price caused by the YMS of airlines that creates the problem of different persons paying such a difference in price for exact the same flight.
2) As also seen in figure 2.1 prices tend to increase when the days left till departure decreases, in the appendix in figure A.1 we have plotted the mean price (red) and the median price (blue) of flights onto the number of days before departure. For all of these flights we can see that both increase when nearing the date of departure. Suggesting that days left till departure becomes more important when nearing the departure date. 3) We would expect that there is a difference between a flight flying on a Wednesday and on a Sunday, flying on a Wednesday should be a regular flight for 'normal' price. While flying on a Sunday would be regarded as a more expensive flight because it flies in the weekend.

2.2 Linear Quantile Mixed Model

Based on the data we have we could try to fit a linear model, to fit such a model we could use in the simple case just a linear line trough the points fitted by ordinary least squares regression as exampled in figure 2.2.

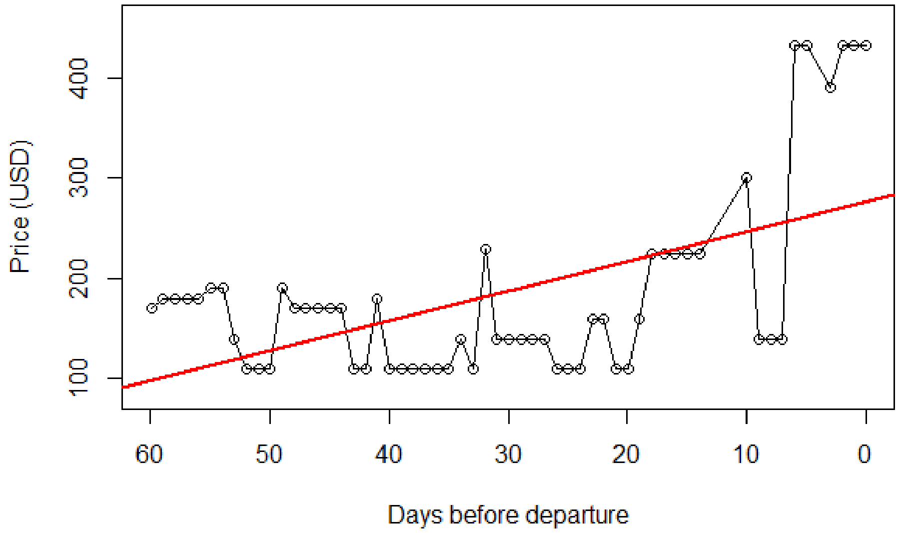

Figure 2.2: Linear fit on flight DL1940 departing from SFO heading to JFK at 2012-03-27

$$y_i = \alpha + \beta x_i + \epsilon_i \tag{2.1}$$

The mathematical description of this linear line is given in equation 2.1 where i indicates observation i. x_i is the number of days left till departure, y_i is the predicted price in USD for this observation and α,β, ϵ are respectively the estimated intercept, slope parameter and residuals. Assumed in this model is that each x_i is independent. The slope parameter β indicates how the price behaves as the days left decreases. This simple model can therefor give us information about the price of this flight based on the number of days left before departure.
This ordinary least squares regression fits the model to the mean, such that this model only tells us about how the mean of the price behaves as number of days till departure decreases. This is not what we and the consumers are interested in, what we would like to know is how the *lowest* prices behave during, say the 60 days period before departure. So instead of a simple linear regression on all the price observations, we would like to only use the low prices for our regression. Therefor we will be using quantile regression on a low quantile as exemplified in figure 2.3 as this would be more suitable for the problem we are try to solve. Selecting a really low quantile τ would give us a model about really low prices, but as we are lowering the τ less price observations becomes relevant. While we would like to predict a really low quantile, the selection of a τ is a trade off between modeling lower prices

and the predictive power of the model as we get fewer data when selecting a lower τ.

We would expect a linear line as in figure 2.2 not to fit the the data quite well, as y_i (the price) does not seem to move linear with a decrease in x_i (the number of days till departure). A logarithmic model could be a possible better fit to the relationships between y_i and x_i as we would expect that the number of days before departure would get more important nearing the day of departure and such it would increase the price exponentially instead of linear. Such a model is given in equation 3.1b

Another model that is interesting to try to fit on this data as well, is a quadratic model as given in equation 2.2. Such a model does have a minimum price, if β_2 is a positive integer, due to its convex form. Then the corresponding minimum price moment, the number of days before departure when the price is at is lowest, can be calculated. Having such a 'optimal' buying day in our model could be proven useful for air travelers wanting to know when to buy the cheapest tickets.

$$y_i = \alpha + \beta_1 log_{10}(x_i + 1) + \beta_2 log_{10}(x_i + 1)^2 + \epsilon_i \tag{2.2}$$

Again y_i is the predicted price in USD and x_i the number of days before departure for observation i. We add 1 to the x_i's such that $log_{10}(x_i + 1)$ is not undefined if $x_i = 0$ and $log_{10}(x_i + 1)$ is positive on the whole domain $(0 \leq x_i \leq 60)$.

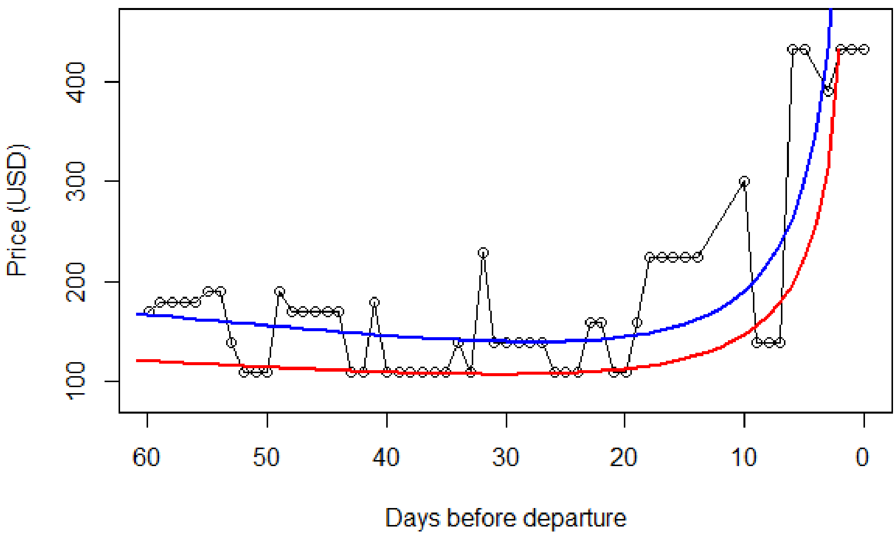

Figure 2.3: Second order polynomial fit on 20^{th} quantile (red) and on the 50^{th} quantile (blue) of flight DL1940 departing from SFO heading to JFK at 2012-03-27

We consider flights with the same flight number, with the difference that it departs on another day, as an independent instance of the same flight. While these flight instances have some same effects called fixed effects, every flight instance has their own effects called random effects. For example, we assume that the influence of the number of days left is fixed, are constant across all flights. But we also assume that the intercepts of these flight instances are random, such that the predictions for the tickets prices correspond to parallel lines. Linear mixed models can capture both random and fixed effects thus generate a model that fits the behavior of the ticket price better (relative to using only fixed effects).

The combination of using linear mixed models while being interested in a low price (read: low quantiles) explains the use of linear quantile mixed models [7] in this research. We will be using the lqmm package in R which is the R-implementation of Linear Quantile Mixed Models by Marco Geraci [6]. We will not be making any modifications to this package, but just use standard implementation as it is capable of fulfilling our needs. The lqmm package fits linear quantile mixed models based on the asymmetric Laplace distribution, as described in [7]. Lqmm requires several parameters as its input, we will be mostly using default parameters but some we will define ourselves, such as the: fixed effects, random effects, grouping factor and τ, which is the quantile to be estimated.

Chapter 3

Results

To select a good model on this data we have to select good explanatory variables which can describe our explained variable y . This model should not have too many variables as it will over fit but also not too few as it will not describe the data properly. From chapter 2.1 we selected the variables: days_left (0-60), as the numbers of days left till departure and is_weekday, which can be weekend or weekday. With these variables we have given several suggestions in chapter 2.2 of how these models could explain y, below these are summarized.

$$y = \alpha + \beta_1 days\_left_i + \epsilon_i \tag{3.1a}$$

$$y = \alpha + \beta_1 log_{10}(days\_left_i + 1) + \epsilon_i \tag{3.1b}$$

$$\begin{aligned} y = \alpha &+ \beta_1 log_{10}(days\_left_i + 1) \\ &+ \beta_2 log_{10}(days\_left_i + 1)^2 + \epsilon_i \end{aligned} \tag{3.1c}$$

$$\begin{aligned} y = \alpha &+ \beta_1 log_{10}(days\_left_i + 1) \\ &+ \beta_2 log_{10}(days\_left_i + 1)^2 \\ &+ \beta_3 is\_weekday_i + \epsilon_i \end{aligned} \tag{3.1d}$$

We have fitted these models on our data for each specific flight from table 2.1 to see how these models fit the data and whether the variable is_weekday is significant. For the selection between the models we will use the Akaike information criterion (AIC), which tells us which model is a relatively better fit (lower is better). Below is given table 3.1 which summarizes the AIC of the fit of each of the models given above for every flight in table 2.1, for $\tau = 0.1$.

Airline	*Flight Number*	*Model 3.1a*	*Model 3.1b*	*Model3.1c*	*Model 3.1d*
American Airlines	AA12	108,038	106,358	106,031	106,027
United Airlines	UA286	329,478	319,920	319,537	319,002
Delta Airlines	DL1940	211,962	206,698	205,697	205,269
JetBlue	B6648	327,252	322,698	320,437	320,498
Virgin America	VX22	357,165	349,974	349,983	349,781
US Airways	US6634	168,122	164,020	164,118	164,008

Table 3.1: The AIC for each of the models given above, for one specific flight of each airline, $\tau = 0.1$

Airline	*Flight Number*	*Intercept*	β_1	β_2	β_3
American Airlines	AA12	499.325***	-200.551***	41.603***	-60.836***
United Airlines	UA286	615.669***	-391.746**	119.860*	-92.104*
Delta Airlines	DL1940	547.310***	-413.333***	138.113***	-54.323*
JetBlue	B6648	481.1708***	-334.3723***	102.8611***	-53.0349***
Virgin America	VX22	474.1582***	-194.7038***	32.9968***	-46.3316***
US Airways	US6634	598.871***	-314.523***	78.046*	-48.639

Table 3.2: The coefficients of the explanatory variables of model 3.1d for all airlines, $\tau = 0.1$, Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1.

Since the AIC for model 3.1d is (mostly) lower than the AIC of model 3.1c such that it is a relatively better fit for our data we will further elaborate on this model. The estimated coefficients of this model are given in table 3.2, from this table we would like to mention four things, firstly we see that for these flights the explanatory variables used in model 3.1d are significant (with exception of the β_3 estimate for flight US6634). For all flights more information about the models can be found in appendix A. Secondly, we see that the intercepts of these flights are fairly high compared to regular ticket prices, for example the estimated intercept of flight AA12 is 499.325 while the mean ticket price for this flight is \$331,- This high intercept explains us that on the day of the departure of this flight, when the number of days left till departure is equal to 0 and parameters β_1, β_2 drop out, the ticket prices are fairly expensive. Thirdly, the estimated parameter β_2 is negative for each of these flights, meaning that model 3.1d is concave for these flights and such have a minimum estimated ticket price. These minimum ticket prices are given in table 3.3 together with the days before departure we expect these prices, note these prices are for weekend tickets else we should have added the estimates for the β_3 coefficient. As last we would like to mention that all signs for the β_3 are negative, indicating that weekday tickets are cheaper than weekend tickets.

Airline	*Flight Number*	*Minimum price*	*Days before departure*
American Airlines	AA12	\$ 257,63	257
United Airlines	UA286	\$ 295.58	43
Delta Airlines	DL1940	\$ 238.06	31
JetBlue	B6648	\$ 209.43	42
Virgin America	VX22	\$ 186.92	892
US Airways	US6634	\$ 281,99	104

Table 3.3: The expected minimum price for 6 weekend flights with the expected days before departure this occurs. According to the prediction of model 3.1d, $\tau = 0.1$

While three of these best moments to buy airline tickets are outside of our window ($\leq$ 60 days before departure) three others are inside: 31,42 and 43 days before departure. As we can only say something about the expected prices within our window we would expect that in general lowest prices would arrive around 31-43 days before departure. Below is given two series of observations with model 3.1d fitted for flight DL1940 (figure 3.1) , one departures at 2012-03-26 and the other one a day later on 2012-03-27.

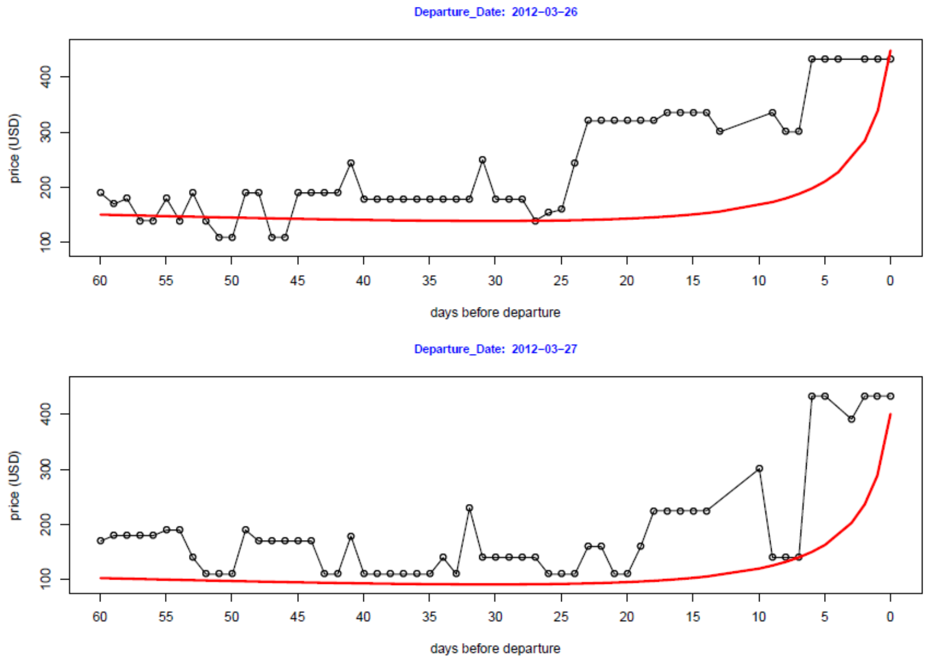

Figure 3.1: Linear quantile mixed model fit for flight DL1940, departure date at 2012-03-26 and at 2012-03-27 ,$\tau = 0.1$.

As we can see in this figure our model follows the movement of the lowest prices for both flights fairly well. It is interesting to see that at the flight leaving at 2012-03-27, 9-7 days before departure there is a price drop, which our model takes into account. It is only in the last days before departure that the predictions of our model starts to deviate from the observations.

Chapter 4

Discussion

The novelty of this research includes the use of quantile regression in combination with mixed models on airline tickets. As there is currently no study about the (low) quantiles of airline prices. Such that we relay the focus from predicting mean ticket prices to predicting the really low prices. The predictors we introduced are good predictors for these 6 flights from different Airlines. With this model more information, such as confidence intervals and std. errors can be given than just a binary buy/wait advice that HAMLETT gives. While we suggested model 3.1d for these flights it is arguable that 3.1c also has an equally good fit on these and other flights.

Chapter 5

Conclusions

This research takes the first step into applying linear quantile mixed model regression on airline tickets, we have shown a possible model for predicting prices of airline tickets based on the number of days left till departure and if the flight leaves in the weekend or on a weekday. Results show that this model follows cheap tickets prices in many days before departure fairly well, but tends not to be very effective several days before departure as it just not quite capture the behavior.

5.1 Future Work

We think several steps can be taken to improve performance, therefor further research is needed. Such as for example the inclusion of more predictors for the price, such as the fuel prices, the distance between airports, holidays and restrictions on a specific flight tickets. In the optimal case the number of seats left would be available but if not, perhaps it is possible to estimate this or derive the ticket demand.
More work can also be done on the applicability of the algorithm, as we only applied it to certain flights from SFO to JFK. Whether to make this algorithm commercially available it has to be converted into an algorithm which can decreasing ticket costs also it has to be applicable to more flights and on different routes. Also the data we used in this research are daily observations of the flights between these two airports, by increasing the number of observations more data becomes available which means more data can be used by the lowest quantiles. More observations means also the possibility to lower the quantile τ, as further research has to show which τ 's are useful for decreasing consumer costs.

5.2 Acknowledgements

We would like to thank Tjeerd Dijksta and Syed Saiden Abbas for providing us with this great subject of applying linear quantile mixed models to airline tickets and for their assistance and guidance during the creation of this thesis. Furthermore we would like to thank Flyr, for providing us with the data this paper is based upon.

Bibliography

[1] Flyr - travel ahead. http://getflyr.com/, August 2014.

[2] Infare solutions. http://www.infare.com/about-us/introduction.aspx, August 2014.

[3] Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.

[4] Jan K. Brueckner, Nichola J. Dyer, and Pablo T. Spiller. Fare determination in airline hub-and-spoke networks. *The RAND Journal of Economics*, 23(3):pp. 309–333, 1992.

[5] Oren Etzioni, Rattapoom Tuchinda, Craig A Knoblock, and Alexander Yates. To buy or not to buy: mining airfare data to minimize ticket purchase price. In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 119–128. ACM, 2003.

[6] Marco Geraci. Linear quantile mixed models: the lqmm package for laplace quantile regression. *J Statist Soft*, 2014.

[7] Marco Geraci and Matteo Bottai. Linear quantile mixed models. *Statistics and computing*, 24(3):461–479, 2014.

[8] William Groves and Maria Gini. A regression model for predicting optimal purchase timing for airline tickets. Technical report, Technical Report 11-025, University of Minnesota, Minneapolis, MN, 2011.

[9] Martin Jacobsen. *Point process theory and applications*. Springer, 2006.

[10] Sheryl E. Kimes. The basics of yield management. *The Cornell Hotel and Restaurant Administration Quarterly*, 30(3):14 – 19, 1989.

[11] Sheryl E. Kimes. Yield management: A tool for capacity-considered service firms. *Journal of Operations Management*, 8(4):348 – 363, 1989.

[12] T. Wohlfarth, S. Clemencon, F. Roueff, and X. Casellato. A data-mining approach to travel price forecasting. In *Machine Learning and Applications and Workshops (ICMLA), 2011 10th International Conference on*, volume 1, pages 84–89, Dec 2011.

Appendix A

Appendix

Below are given summaries of the regression, including AIC, Std. Error and lower/upper bound. These regressions are all on the 10^{th} quantile.

```
==========================Flight AA12==========================
Call: lqmm(fixed = price_inc ~ days_left, random = ~1,
group = outbound_departure_date, tau = taus, data = dd)

Fixed effects:
                Value Std. Error lower bound upper bound  Pr(>|t|)
(Intercept) 398.48113    8.13976   382.12367    414.8386 < 2.2e-16 ***
days_left    -2.00000    0.39791    -2.79962     -1.2004 7.068e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Null model (likelihood ratio):
[1] 5726 (p = 0)
AIC:
[1] 108038 (df = 4)
---
Call: lqmm(fixed = price_inc ~ logDays, random = ~1,
group = outbound_departure_date, tau = taus, data = dd)

Fixed effects:
               Value Std. Error lower bound upper bound  Pr(>|t|)
(Intercept)  398.365     25.950     346.218     450.513 < 2.2e-16 ***
logDays     -107.348     15.213    -137.919     -76.777 5.424e-09 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Null model (likelihood ratio):
[1] 7406 (p = 0)
AIC:
[1] 106358 (df = 4)
```

```
Call: lqmm(fixed = price_inc ~ logDays + logDaysSqr, random = ~1,
group = outbound_departure_date, tau = taus, data = dd)

Fixed effects:
               Value Std. Error lower bound upper bound  Pr(>|t|)
(Intercept)  446.007     16.812     412.222     479.792 < 2.2e-16 ***
logDays     -207.496     20.825    -249.346    -165.647  2.27e-13 ***
logDaysSqr    44.124     12.017      19.975      68.274 0.0005951 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Null model (likelihood ratio):
[1] 7735 (p = 0)
AIC:
[1] 106031 (df = 5)
Call: lqmm(fixed = price_inc ~ logDays + logDaysSqr + is_weekend, random = ~1,
    group = outbound_departure_date, tau = taus, data = dd)

Fixed effects:
                     Value Std. Error lower bound upper bound  Pr(>|t|)
(Intercept)        499.325     12.605     473.995     524.656 < 2.2e-16 ***
logDays           -200.551     17.900    -236.523    -164.580 4.045e-15 ***
logDaysSqr          41.603     10.544      20.415      62.791 0.0002535 ***
is_weekendweekday  -60.836     14.552     -90.080     -31.593 0.0001196 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Null model (likelihood ratio):
[1] 7742 (p = 0)
AIC:
[1] 106027 (df = 6)
```

```
==========================Flight UA286==========================
Call: lqmm(fixed = price_inc ~ days_left, random = ~1,
group = outbound_departure_date, tau = taus, data = dd)

Fixed effects:
               Value Std. Error lower bound upper bound  Pr(>|t|)
(Intercept) 502.9345    59.1251    384.1181    621.7508 3.248e-11 ***
days_left    -1.9850     0.3826     -2.7538     -1.2161 4.046e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Null model (likelihood ratio):
[1] 10358 (p = 0)
AIC:
[1] 329478 (df = 4)
---
                Value Std. Error lower bound upper bound  Pr(>|t|)
(Intercept)  396.2855    16.8011    362.5224     430.048 < 2.2e-16 ***
logDays     -109.7643     9.9831   -129.8261     -89.702 7.868e-15 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Null model (likelihood ratio):
[1] 19915 (p = 0)
AIC:
[1] 319920 (df = 4)
Call: lqmm(fixed = price_inc ~ logDays + logDaysSqr, random = ~1,
group = outbound_departure_date, tau = taus, data = dd)

Fixed effects:
               Value Std. Error lower bound upper bound  Pr(>|t|)
(Intercept)  616.300     69.756     476.119      756.48 1.040e-11 ***
logDays     -480.540    108.684    -698.949     -262.13 5.439e-05 ***
logDaysSqr   154.485     44.084      65.896      243.07 0.0009883 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Null model (likelihood ratio):
[1] 20300 (p = 0)
AIC:
[1] 319537 (df = 5)
Call: lqmm(fixed = price_inc ~ logDays + logDaysSqr + is_weekend, random = ~1,
    group = outbound_departure_date, tau = taus, data = dd)

Fixed effects:
                     Value Std. Error lower bound upper bound  Pr(>|t|)
(Intercept)        615.669     86.826     441.186    790.1517 4.798e-09 ***
```

```
logDays           -391.746    116.227    -625.313   -158.1802   0.00147 **
logDaysSqr         119.860     45.478      28.469    211.2505   0.01122 *
is_weekendweekday  -92.104     40.929    -174.353     -9.8549   0.02895 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Null model (likelihood ratio):
[1] 20837 (p = 0)
AIC:
[1] 319002 (df = 6)
```

```
==========================Flight DL1940==========================
Call: lqmm(fixed = price_inc ~ days_left, random = ~1,
group = outbound_departure_date, tau = taus, data = dd)

Fixed effects:
                Value Std. Error lower bound upper bound  Pr(>|t|)
(Intercept) 348.21051   35.76134   276.34540    420.0756 4.839e-13 ***
days_left    -1.75099    0.46536    -2.68617     -0.8158 0.0004498 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Null model (likelihood ratio):
[1] 6934 (p = 0)
AIC:
[1] 211962 (df = 4)
---
Call: lqmm(fixed = price_inc ~ logDays, random = ~1,
 group = outbound_departure_date, tau = taus, data = dd)

Fixed effects:
               Value Std. Error lower bound upper bound  Pr(>|t|)
(Intercept)  286.590     19.852     246.695     326.486 < 2.2e-16 ***
logDays      -64.748     18.781    -102.490     -27.006  0.001171 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Null model (likelihood ratio):
[1] 12197 (p = 0)
AIC:
[1] 206698 (df = 4)
Call: lqmm(fixed = price_inc ~ logDays + logDaysSqr, random = ~1,
group = outbound_departure_date,  tau = taus, data = dd)

Fixed effects:
               Value Std. Error lower bound upper bound  Pr(>|t|)
(Intercept)  539.208     25.773     487.416      591.00 < 2.2e-16 ***
logDays     -442.240     29.050    -500.618     -383.86 < 2.2e-16 ***
logDaysSqr   147.466     12.520     122.306      172.63 6.691e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Null model (likelihood ratio):
[1] 13201 (p = 0)
AIC:
[1] 205697 (df = 5)
Call: lqmm(fixed = price_inc ~ logDays + logDaysSqr + is_weekend, random = ~1,
    group = outbound_departure_date, tau = taus, data = dd)
```

```
Fixed effects:
                      Value Std. Error lower bound upper bound Pr(>|t|)
(Intercept)         547.310     26.131     494.799     599.822  < 2e-16 ***
logDays            -413.333     22.909    -459.370    -367.297  < 2e-16 ***
logDaysSqr          138.113     10.725     116.561     159.666  < 2e-16 ***
is_weekendweekday   -54.323     20.465     -95.449     -13.197  0.01068 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Null model (likelihood ratio):
[1] 13630 (p = 0)
AIC:
[1] 205269 (df = 6)
```

```
==========================Flight B6648==========================
Call: lqmm(fixed = price_inc ~ days_left, random = ~1,
group = outbound_departure_date, tau = taus, data = dd)

Fixed effects:
                Value Std. Error lower bound upper bound  Pr(>|t|)
(Intercept) 318.70680   39.13414   240.06380    397.3498 1.152e-10 ***
days_left    -1.03613    0.45436    -1.94920     -0.1231   0.02697 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Null model (likelihood ratio):
[1] 12558 (p = 0)
AIC:
[1] 327252 (df = 4)
---
Call: lqmm(fixed = price_inc ~ logDays, random = ~1,
group = outbound_departure_date,  tau = taus, data = dd)

Fixed effects:
               Value Std. Error lower bound upper bound  Pr(>|t|)
(Intercept)  323.235     20.340     282.361     364.109 < 2.2e-16 ***
logDays      -95.656     10.676    -117.111     -74.202 6.767e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Null model (likelihood ratio):
[1] 17112 (p = 0)
AIC:
[1] 322698 (df = 4)
Call: lqmm(fixed = price_inc ~ logDays + logDaysSqr, random = ~1,
group = outbound_departure_date, tau = taus, data = dd)

Fixed effects:
                Value Std. Error lower bound upper bound  Pr(>|t|)
(Intercept)  477.4814     7.9147    461.5762      493.39 < 2.2e-16 ***
logDays     -330.7288     7.5282   -345.8573     -315.60 < 2.2e-16 ***
logDaysSqr   102.1448     4.4067     93.2891      111.00 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Null model (likelihood ratio):
[1] 19375 (p = 0)
AIC:
[1] 320437 (df = 5)
Call: lqmm(fixed = price_inc ~ logDays + logDaysSqr + is_weekend, random = ~1,
    group = outbound_departure_date, tau = taus, data = dd)
```

```
Fixed effects:
                      Value Std. Error lower bound upper bound  Pr(>|t|)
(Intercept)        481.1708     6.6536    467.7999     494.542 < 2.2e-16 ***
logDays           -334.3723    10.2843   -355.0394    -313.705 < 2.2e-16 ***
logDaysSqr         102.8611     6.1048     90.5930     115.129 < 2.2e-16 ***
is_weekendweekday  -53.0349     8.5695    -70.2560     -35.814 1.197e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Null model (likelihood ratio):
[1] 19316 (p = 0)
AIC:
[1] 320498 (df = 6)
```

```
==========================Flight VX22==========================
Call: lqmm(fixed = price_inc ~ days_left, random = ~1,
group = outbound_departure_date, tau = taus, data = dd)

Fixed effects:
                Value Std. Error lower bound upper bound  Pr(>|t|)
(Intercept) 416.34923   12.55158   391.12589    441.5726 < 2.2e-16 ***
days_left    -2.62860    0.18633    -3.00304     -2.2542 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Null model (likelihood ratio):
[1] 26075 (p = 0)
AIC:
[1] 357165 (df = 4)
---
Call: lqmm(fixed = price_inc ~ logDays, random = ~1,
group = outbound_departure_date, tau = taus, data = dd)

Fixed effects:
                Value Std. Error lower bound upper bound  Pr(>|t|)
(Intercept)  424.5994     9.9058    404.6929      444.51 < 2.2e-16 ***
logDays     -125.1194     6.5935   -138.3696     -111.87 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Null model (likelihood ratio):
[1] 33266 (p = 0)
AIC:
[1] 349974 (df = 4)
Call: lqmm(fixed = price_inc ~ logDays + logDaysSqr, random = ~1,
group = outbound_departure_date, tau = taus, data = dd)

Fixed effects:
                Value Std. Error lower bound upper bound  Pr(>|t|)
(Intercept)  450.8447    11.4357    427.8637     473.826 < 2.2e-16 ***
logDays     -204.0612    10.8934   -225.9524    -182.170 < 2.2e-16 ***
logDaysSqr    33.7808     7.2485     19.2143      48.347 2.453e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Null model (likelihood ratio):
[1] 33259 (p = 0)
AIC:
[1] 349983 (df = 5)
Call: lqmm(fixed = price_inc ~ logDays + logDaysSqr + is_weekend, random = ~1,
    group = outbound_departure_date, tau = taus, data = dd)
```

```
Fixed effects:
                      Value Std. Error lower bound upper bound  Pr(>|t|)
(Intercept)        474.1582    15.3645    443.2821     505.034 < 2.2e-16 ***
logDays           -194.7038     9.9451   -214.6892    -174.718 < 2.2e-16 ***
logDaysSqr          32.9968     5.4520     22.0405      43.953 1.945e-07 ***
is_weekendweekday  -46.3316    12.7229    -71.8993     -20.764 0.0006526 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Null model (likelihood ratio):
[1] 33463 (p = 0)
AIC:
[1] 349781 (df = 6)
```

```
==========================Flight US6634==========================
Call: lqmm(fixed = price_inc ~ days_left, random = ~1,
group = outbound_departure_date, tau = taus, data = dd)

Fixed effects:
                Value Std. Error lower bound upper bound  Pr(>|t|)
(Intercept) 634.86722   43.81388   546.81993    722.9145 < 2.2e-16 ***
days_left    -4.30002    0.42967    -5.16347     -3.4366 1.961e-13 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Null model (likelihood ratio):
[1] 2514 (p = 0)
AIC:
[1] 168122 (df = 4)
Call: lqmm(fixed = price_inc ~ logDays, random = ~1,
 group = outbound_departure_date, tau = taus, data = dd)

Fixed effects:
                Value Std. Error lower bound upper bound  Pr(>|t|)
(Intercept)  519.2870    35.8029    447.3384      591.24 < 2.2e-16 ***
logDays     -161.4610     6.4315   -174.3855     -148.54 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Null model (likelihood ratio):
[1] 6616 (p = 0)
AIC:
[1] 164020 (df = 4)
Call: lqmm(fixed = price_inc ~ logDays + logDaysSqr, random = ~1,
group = outbound_departure_date, tau = taus, data = dd)

Fixed effects:
               Value Std. Error lower bound upper bound  Pr(>|t|)
(Intercept)  593.914     54.730     483.930      703.90 1.245e-14 ***
logDays     -320.212    101.076    -523.333     -117.09  0.002643 **
logDaysSqr    80.265     47.237     -14.662      175.19  0.095625 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Null model (likelihood ratio):
[1] 6520 (p = 0)
AIC:
[1] 164118 (df = 5)
Call: lqmm(fixed = price_inc ~ logDays + logDaysSqr + is_weekend, random = ~1,
    group = outbound_departure_date, tau = taus, data = dd)
```

```
Fixed effects:
                      Value Std. Error lower bound upper bound  Pr(>|t|)
(Intercept)         598.871     62.922     472.424     725.318 1.012e-12 ***
logDays            -314.523     59.220    -433.529    -195.516 2.640e-06 ***
logDaysSqr           78.046     29.669      18.423     137.669   0.01136 *
is_weekendweekday   -48.639     51.024    -151.175      53.898   0.34514
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Null model (likelihood ratio):
[1] 6632 (p = 0)
AIC:
[1] 164008 (df = 6)
```

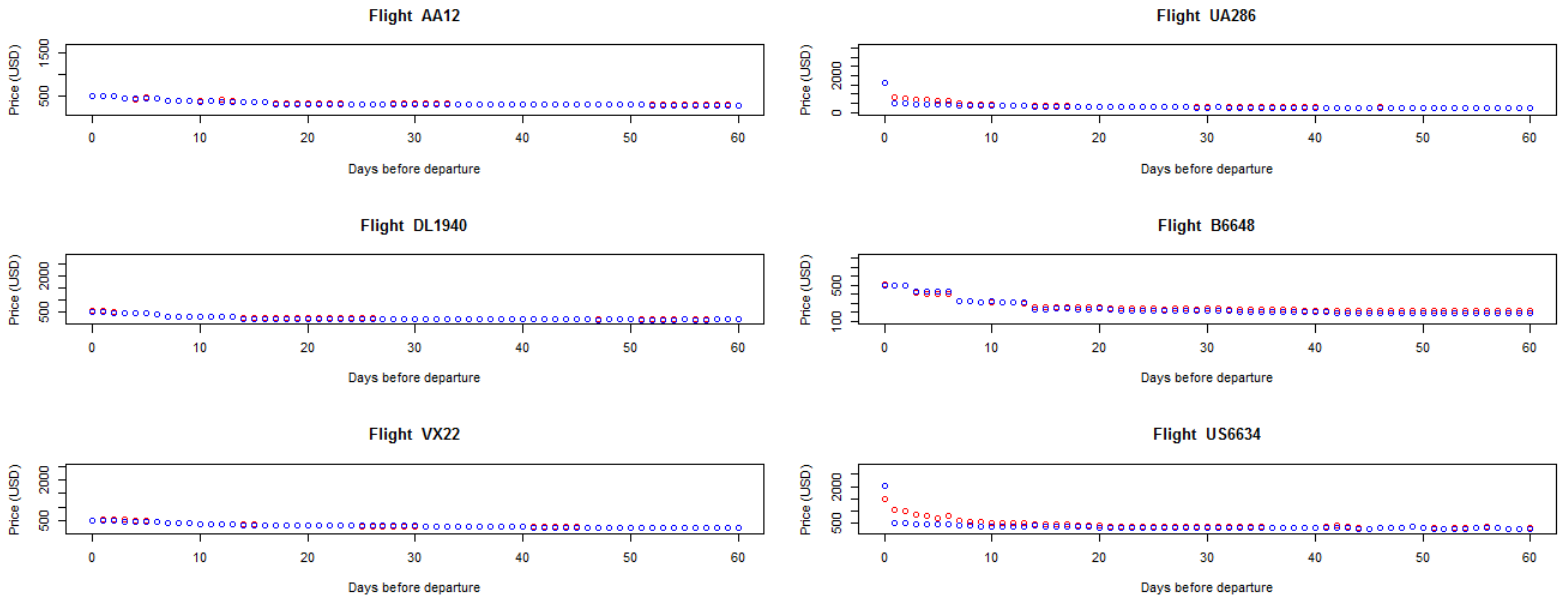

Figure A.1: The mean (red) and median (blue) prices for our six different flights, plotted agains the number of days left.