Procedia Computer Science

Volume 66, 2015, Pages 267–276

YSC 2015. 4th International Young Scientists Conference on Computational Science

Data-driven Modeling of Airlines Pricing

Anastasia Lantseva, Ksenia Mukhina, Anna Nikishova, Sergey Ivanov and Konstantin Knyazkov

ITMO University, Saint-Petersburg, Russia

stasja16@inbox.ru, mukhinaks@gmail.com, annanikishova@gmail.com, sergei.v.ivanov@gmail.com, constantinvk@gmail.com

Abstract

The popularity of travelling by airplanes is constantly growing. Much of existing research describe the global flight market. At the same time, Russian air market is characterized by its peculiarities that have to be identified to build proper models of airfare. The objective of this study is to analyze Russian air transportation market and compare the behavior of prices on local and global flights. Using these data, collected from two independent ticket price information aggregators (AviaSales and Sabre) for the period of spring-summer 2015, an empirical data-driven model was built for air prices prediction for different flight directions. We found that the form of price dependency on purchase earliness differs dramatically between local and international flights in two largest Russian cities (Moscow and Saint-Petersburg).

Keywords: airlines, airlines pricing, Russian airline market, pricing modeling

1 Introduction

Airline ticket price depends on many different dynamic factors, such as airline pricing policies, flight distance, class of service, airline, global population mobility, all of which define the travel demand. Ticket costs can vary significantly for the same flight, even for nearby seats. The model of airline tickets market varies from country to country and depends on the volume and structure of supply (number of airlines and air flights) and demand (number of passengers, seasonal peculiarities).

Air travel demand is affected by the global population mobility, which itself is influenced by various factors. Most people tend to work on the schedule. So, tickets with dates of departure/arrival on the weekend are usually more expensive. Also, the increase in prices can be seen during public holidays. Traditionally, there is a believe that the plane tickets are cheaper if they are bought in advance. Today, this is true only in part. For example, due to the pricing policy airlines offer cheap flights shortly before departure to maximize the load of their flights. It happens due to the sharpening of competition between airlines. The cost also can depend on the seasonal factor. Traditionally, there are three seasonal indicators. High season is characterized by the highest ticket prices (usually

Selection and peer-review under responsibility of the Scientific Programme Committee of YSC 2015

doi:10.1016/j.procs.2015.11.032

December–January, June–September), shoulder season (September–December, March–June) by average prices, and low season by the lowest prices (January–March). Tickets for the low season are always cheaper because people fly on rare occasions, and usually that is the time when airlines organize promotions and sell tickets at discounted prices.

The service type is another main factor affecting ticket price. There are three basic classes of different price categories: economy, business, and first class. Flights can also be divided into regular, charter and low-cost. The latter two types are nearly always the cheapest, but they have their distinctiveness, e.g. in Russia applying low-cost marketing policy is exceedingly rare among airlines. Information on airline pricing policies can be publicly available (e.g. in the form of promotions) or hidden from customers. Promotions often are not made public in advance, so it is hard to plan the purchase of a ticket for the promotion. Usually, there are two windows of time associated with a promotion. Each special offer includes flights, which have very limited number of tickets. Some airlines organize their promotions six months prior to the date of departure while other sell tickets with great success during the season of departure. Airlines want to maximize their profits balancing between filling their flights and increasing prices. The dates of seasons may vary slightly for different airlines.

All the highlighted factors are involved in the price formation process and should be taken into account in designing of personal ticket buying support system. The aim of this research is to create a model, which could be used for airline tickets price forecasting for Russian cities. The contribution of this work includes the following:

- collection, processing and fusion of data from two independent sources on ticket prices for flights from two largest Russian cities;
- analysis of macroscopic trends in price formation, identification of patterns in price formation;
- the data-driven model on the base of collected dataset.

2 Related Works

With the expansion of the Internet around the globe, the ticket selling industry underwent a series of changes (Klein & Loebbecke, 2000). Tickets are available for purchase 24/7 from anywhere in the world, and the potential buyer can compare prices from the great variety of airlines before making his final decision. Due to this fact prediction of the ticket price is becoming a more difficult problem as it depends on many factors. Nevertheless, some of these criteria were studied rather well. For instance, there is a correlation between airline prices and air traffic delays (Forbes, 2008). Authors found out that each one-minute flight delay leads to decrease in the ticket price of about $1.5.

Influence of low-cost airlines on ticket market was studied in some works. The research (Hofer, Windle, & Dresner, 2008) is devoted to the study of results of competition between low-cost airlines and regular service airlines. The advantages of EasyJet, Ryanair are reviewed in (Gillen & Lall, 2004). Work (Barrett, 2004) examines the demand function for airlines in Europe. Moreover, the influence of low-cost airlines on flights and the number of passengers in the European airline market has been impressive. The influence of low-cost airlines on two airports in Europe was studied in the work (Francis, Fidato, & Humphreys, 2003).

Comparison between European and Asian market was made in the paper (O'Connell & Williams, 2005) devoted to the perception of low-cost airlines and airlines with a full set of services. It was found that low-cost carriers attracted predominantly younger people and mostly for business trips. There was no significant difference found in the attitude and perception of passengers from two very different continents. However, there were certain distinctions in price formation.

The empirical investigation of the US market, for American Airlines and United Airlines, in particular, was made in the paper (Brander & Zhang, 1990). For studying oligopoly on the air market

the Bertrand and Cournot competition models were chosen. The welfare effects of fare discrimination across periods in time were measured in the paper (Lazarev, 2013). Lazarev developed an empirical model of optimal prices and found out that due to intertemporal price discrimination airlines can earn approximately 90% of the profit. Pricing behavior of low-cost airlines in the UK was studied in the article (Pitfield, 2005). This research showed that it is reasonable to base a model for analysis on an ARIMA approach. Another implementation of ARIMA model was used in work (Pitfield, 1993) for prediction of air transport demand.

However, the most important question for the general public is "when should someone buy a ticket to minimize the price?". It seems to be reasonable to buy a ticket for the flight as soon as possible. However, this strategy appears not always to be the best option. Groves and Gini's study (Groves & Gini, 2013) shows that it is true only for low-cost airlines in case of traveling across the USA. They introduced an agent who was able to optimize a purchase. In the paper (Etzioni, 2003) the multi-strategy data mining algorithm was presented to find the optimal time for purchasing (for the last 21 days prior to departure). The price behavior before the flight in the London–Paris market was analyzed in work (Pels & Rietveld, 2004). They examined the lowest available price per carrier for four months. After statistical analysis, it was found that there is the upward trend of the cost over time for some airlines, whereas other companies retain almost constant price for the whole period.

3 Methods

Departure cities selection. Considering a large number of investigations in the field of airline markets in Europe, Asia, and the USA, it seems reasonable to investigate the Russian air market separately. There are two major hubs in Russia: Moscow and Saint-Petersburg. They have an impressive part of the whole flights number across the country. Four largest Russian airports serve almost 100 million passengers per year (Moscow Domodedovo Airport), (Sheremetyevo International Airport), (Vnukovo International Airport, JSC.) and (Northern Capital Gateway LLC). Three airports are located in Moscow and one in Saint-Petersburg. Also, there is a lack of intercity flights in Russia, so people who want to reach their destinations, have to choose flights with a change in Moscow or Saint-Petersburg. Obviously, it has some impact on price formation since airplanes on these routes have no empty seats under almost any circumstances. Due to these facts, the optimal date for ticket purchasing for these cities should be studied independently.

Destinations sampling. For this study, these cities were selected as cities of departure. For these points of departure, a number of destinations were chosen. Destination sampling was performed based on flights popularity taking into account constraints of data sources. The initial sample was fetched via AviaSales API, which allows requesting 50 most popular international flights from Moscow. Same directions were taken for Saint-Petersburg (further this group of destinations is named "World"). In figure 1 all destinations are presented. International flights group is characterized by the domination of European cities. Local destinations include 40 largest cities of the Russia.

For each day ("date of purchase") a request to get the minimum price has been done for the flight on each day. The collection period is 75 days for AviaSales and 90 days for Saber for each direction. Since the data fetched from datasets have distinctions, we use a single dataset to build hypotheses and to identify specific patterns, and use another one to verify the results and to build a pricing model.

Methods of analysis. The collected dataset has the following dimensions: city of departure, destination, ticket purchase date, departure date, ticket options with the price. The price heat maps are chosen as the main tool for analysis. They are built for fixed departure city and destination. The data in this form demonstrates the price for each day, and visual analysis can be easily conducted.

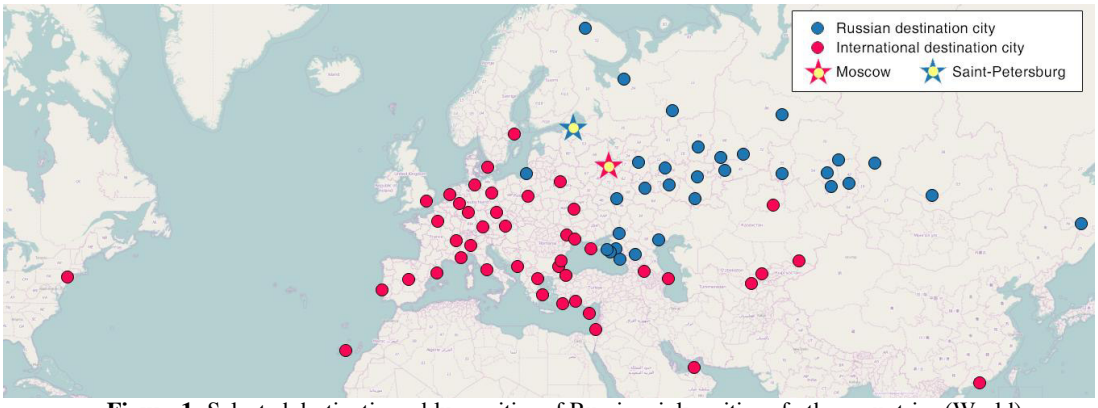

Figure 1: Selected destinations: blue – cities of Russia, pink – cities of other countries (World)

3.1 Data acquisition

To be able to validate the collected data, two independent data sources were used: Sabre Services (Sabre, 2015) and AviaSales (GO TRAVEL UN LIMITED, 2015). In both cases, corresponding public APIs were used. The obtained data includes full itinerary (the IATA-codes of departure and destination), a distance between cities, departure and return dates, and carrier operating the flight, ticket price, currency rates and number of transfers. It is necessary to note that the data is incomplete in both datasets and contain gaps caused by technical problems.

Data source 1. AviaSales is a meta-search engine for tickets containing information about 728 world airlines, 45 agencies, and five booking systems. The strength of this data source is that this search engine provides data for all selected destinations. However, the specifics of the requests was that only the minimum price for each day was available.

Data source 2. Sabre (Sabre Travel Network) is a part of the group of companies Sabre Holdings (USA), serving travel agencies, travel suppliers, corporations and government agencies in 59 countries around the world. Due to the business model of Sabre, it works only with the limited number of airlines. So, is does not provide information on all selected destinations.

The key differences between two datasets are listed in the Table 1.

Data source:	AviaSales	Sabre
Stored flight option prices:	1 (minimum)	≤300
Dates of collection:	03.04.2015–03.06.2015	06.04.2015–14.08.2015
Number of directions (total):	90	41
– from Moscow to World:	50	33
– from Saint-Petersburg to World:	50	2
– from Moscow to Russia:	40	8
– from Saint-Petersburg to Russia:	40	2

Table 1: Comparison of data sets

4 Results analysis

In Figure 2 the aggregated data on price trends is presented for different cities of departure and groups of destination cities. After normalization of all directions all the data aggregated together by computing price per kilometer for several destinations. The color of heatmaps for different directions and data sources are shown in Figure 2.

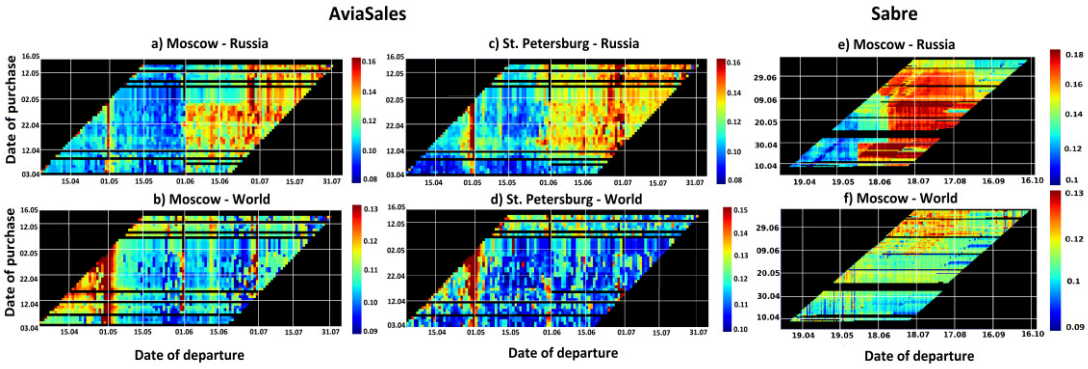

Figure 2: Behavior of the price per kilometer depending on purchase date for both datasets (scale is in EUR per km): a) from Moscow to Russia, b) from Moscow to World, c) from St. Petersburg to Russia, d) from St. Petersburg to World (all AviaSales), e) from Moscow to Russia (Sabre) and f) from Moscow to World (Sabre)

In these figures, vertical line patterns are visible, which correspond to causes of price change specific to particular date or season. It means that the price for an appropriate date of departure will be high/low regardless of the date of purchase. For example, the rise of prices for the holiday on May 1 is visible from all directions. For this holiday it can be seen that earliness of purchase affects the price: the closer to diagonal the wider the red band is. The same vertical pattern appears with another intensity on June 12. The seasonal price increase is clearly visible for flights across Russia. It starts on July 1 and lasts for months. This observation can hardly be explained by customers' demand, as it has a very sharp left border. Moreover, this seasonal shift in prices has also sharp upper border (it is visible for flights from Moscow on May 1), which make its shape look like the step. Here we can see a horizontal pattern, which is connected to particular purchase date and also can be hardly explained with demand specific. More likely it has reasons relating to airfare policy specific for summertime season. The same patterns are also visible in Dataset 2 (see figure 2e). The "step" behavior is reproduced after that once more: left border on July 1, upper border close to June 1 (Figure 2e). Diagonal patterns can be seen in figures corresponding to World direction: the closer ticket purchase to flight date, the higher price. Also, several diagonal patterns can be observed in the second dataset. Prices decreased about two weeks before the flight are seen in Figure 2e. Prices increased with the same time gap can be seen in Figure 2f.

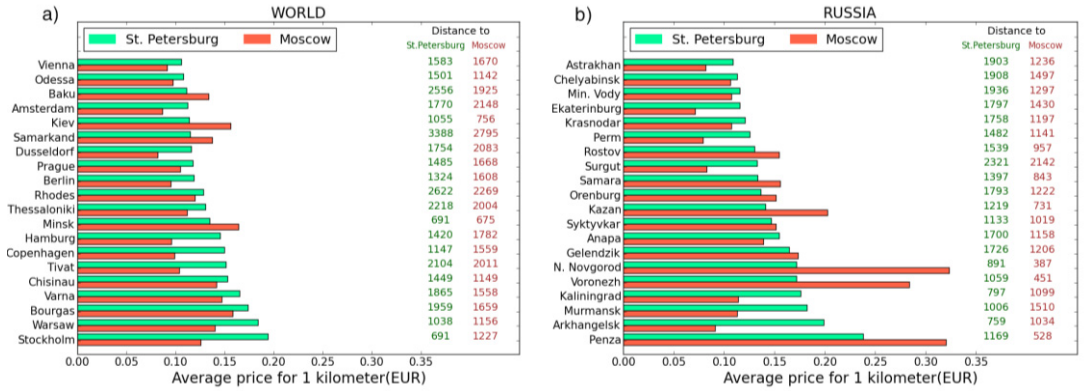

Figure 3: Top 20 most expensive cities by the cost per kilometer (sorted for Saint-Petersburg): a) global flights; b) local flights

It should be noted that results for two different departure cities agrees to each other but intensity of patterns of price change for Saint-Petersburg is lower than for Moscow. It can be explained by the

following: a sample of flights from Saint-Petersburg is less and also larger amount of flights is performed via Moscow.

The second dataset contains a slightly fewer number of directions but has a larger range of dates (Figure 2e, Figure 2f). The results obtained from it agree with the first dataset in overall, but there are some peculiarities, e.g. some diagonal patterns are seen only in Sabre data.

Figure 3 shows the cost price per kilometer in each direction with a particular departure city for 20 locations with the highest price per kilometer for flights departing from Saint-Petersburg. One can see that relative order for different directions is not the same for two departure cities. It can occur due to complex price formation: ticket price contains fixed part, which doesn't depend on distance.

Figure 4 shows the average cost per kilometer for local and international flights as a function of ticket purchase earliness. The curves of the price dependency on purchase earliness for international flights and domestic flights differ dramatically. The curve corresponding to international flights decreases monotonically up to some minimum; then the price increases insignificantly. For the flights inside Russia the curve has opposite shape: it increases with increasing of the time interval before departure. This rise of flights prices inside Russia can be explained by the influence of the seasonal airfare policies. So, for flights to Russia season is the most important criteria, and the purchase in advance does not give the desired effect. However, the early purchase can be profitable for international flights.

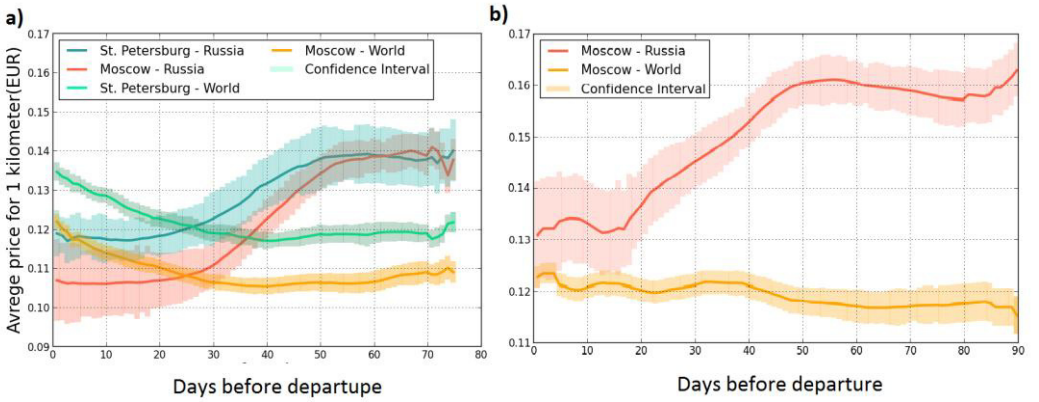

Figure 4. Average cost per kilometer of local and global flights a) from Moscow and St. Petersburg, taking into account prematurity of purchase up to 75 days b) from Moscow and up to 90 days

5 Prediction of Ticket Prices with Regression Model

To get the price prediction model, the analysis results obtained from dataset 1 were used. For this purpose, two following approximations are proposed. The collected data for the local flights from the both departure cities (Figure 5a, Figure 5b) can be approximated as a sigmoidal function described by the formula:

$$f(x,a,b,c,d) = a + b \cdot \left(1 + \exp\left(\frac{-(x-c)}{d}\right)\right)^{-1}, \quad (1)$$

where f gives cost per kilometer, x is number of days before departure, a – initial price, b – maximum rate of price, c – the inflection point of sigmoid, d – coefficient of curvature.

The same dependency for international flights can be approximated with cubic parabola (Figure 5c, Figure 5d):

$$f(x,a',b',c',d') = a' + b' \cdot x + c' \cdot x^2 + d' \cdot x^3, \quad (2)$$

where f gives cost per kilometer, x – number of days departure, a' – initial price, b' – coefficient of curvature, c' – pitch angle of parabola, d' – coefficient of growth trend (fall).

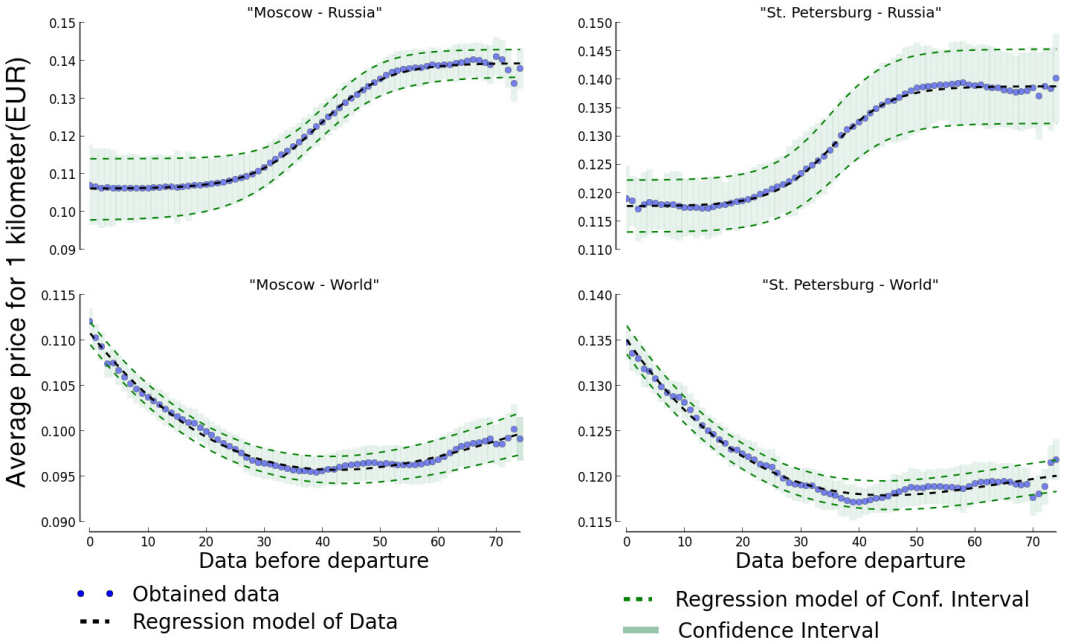

Figure 5. Discrete values of the average price per kilometer along with the 95% confidence interval and regression model

For each of the directions coefficients of the models were determined. Results are presented in Table 2. These coefficients determine the behavior of the curve and its trends. The confidence intervals provide an opportunity to build a forecast of the price dispersion in the period considered (whole 75 days prior to departure).

Destination	Local			World	
Departure	Moscow	St. Petersburg		Moscow	St. Petersburg
Average price:					
a	1.06e-1	1.17e-1	a'	1.1e-1	1.3e-1
b	3.32e-2	2.11e-2	b'	- 8.1e-4	- 9.2e-4
c	39.26	35.37	c'	1.3e-5	1.5e-5
d	5.715	5.08	d'	- 5.8e-8	- 8.05e-8
Upper bound of confidence interval:					
a	1.14e-1	1.22e-1	a'	1.1e-1	1.3e-1
b	2.89e-2	2.31e-2	b'	-8.1e-4	-9.4e-4
c	40.72	34.97	c'	1.3e-5	1.6e-5
d	4.889	4.992	d'	-5.83e-8	-8.5e-8
Lower bound of confidence interval:					
a	9.77e-2	9.77e-2	a'	1.09e-1	1.3e-1
b	3.8e-2	3.8e-2	b'	-8.09e-4	- 9.07e-4
c	37.71	37.71	c'	1.3e-5	1.51e-5
d	6.589	6.589	d'	-5.91e-8	-7.6e-8

Table 2: Coefficients of the regression model for the global and local flights

The regression model and more precisely, the equations 1 and 2 predict the behavior of rates during 90 days before departure. It allows to forecast behavior values of price, with the help of comparison of confidence intervals. When choosing the period of flight, one should focus on the

expected confidence intervals and estimate the period. For this purpose, there are criteria for assessing the overlap of confidence intervals. This criterion is based on a comparison of these two confidence intervals, to determine the best time for purchase of ticket in advance: $ci(x_k)$ - confidence interval, where $x_k \in X$ (days before depature) $k \in [0,90]$;

- $ci(x_i) \cap ci(x_j) = 0, i \neq j$ – high degree of price volatility;
- $ci(x_i) \cap ci(x_j) < 50\%, i \neq j$ – low level of confidence;
- $ci(x_i) \cap ci(x_j) \geq 50\%, i \neq j$ – price stability

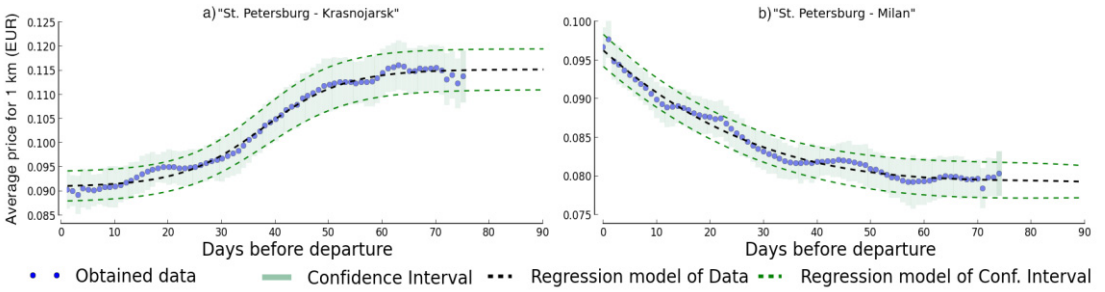

Figure 6. Result of regression model with period at 90 days for two directions: a) St. Petersburg – Krasnoyarsk b) Petersburg – Milan.

If the overlap of confidence intervals is more than 50%, it means the price will stay stable for a while, and this is a good time to buy. As an example, we consider a local flight from St. Petersburg to Krasnoyarsk (see Figure 6a). Using the proposed criteria, we can say that for the domestic flights acceptable period of purchase is up to 40 days before departure. The period shows stable behavior and has no overlap with a period after the 50th day, which means the price is likely to change much after the 40th day. Comparisons of confidence intervals indicate instability of prices in the period from 40th to 50th day; here it is impossible to predict the behavior of prices.

The same analysis is presented for international flight St. Petersburg – Milan (Figure 6b), here is the opposite trend in the price. In the period prior to the 20th day and in the period 40-90 days, the price is stable. However, the overlap of confidence intervals tends to be zero. It indicates a significant change in the price after 20th day.

6 Discussion and Future Work

Two different markets have been studied in this paper. The first is international flights with the domination of European directions, where a policy of air travel is sufficiently developed, and the second is regional transportation across Russia. Some significant differences were found. On the international flights, where foreign airlines are presented rather well, observed pricing behavior is quite close to that discussed in previous studies. Early ticket purchasing has an advantage over deals at the last moment (Groves & Gini, 2013). Nevertheless, in case of local flights, the situation became controversial and required additional investigations. This is perhaps because competition on the Russian market of air travel is not as high as on international markets. The development of this infrastructure is lagging behind in comparison with foreign countries. Popular in the world “low-cost” carriers have long been an integral part of international flights, however for Russia it remains a novelty. Also, monopolies play an important role. They are very common among Russian airlines such as “NorthStar”, the trademark of airline “Taimyr”, which performs charter and regular flights from Krasnoyarsk, Norilsk, Novosibirsk and other Siberian cities with a focus on flights to Siberian Federal

District. Moreover, strong connection between price per kilometer and flight distance was discovered; for longer distances price tends to be sufficiently lower.

The model which is aimed to be able to predict the prices for the specific dates and to find an optimal date for ticket purchasing has been proposed. The model showed that for international flights it is better to buy a ticket as soon as possible and for the local flight situation can go both ways. However, the 40th day before the departure seems acceptable for purchasing in both directions. At this time, the chances to buy tickets profitably are maximal.

However, performed data analysis has some limitations. Obtained data is incomplete and cannot represent the situation on the whole market. A short period of data collection makes the proposed models relevant for the corresponding time interval. It is especially important for directions across Russia, as available data reflects seasonal peculiarities that should be explained. There are two directions to tackle this issue: the first one is to expand the period of data collection (a year or more). It will allow us to trace pricing behavior not only for the spring-summer season. The second one is to extend data collected for the each flight, which will allow to observe a behavior of each particular airline during ticket selling.

7 Conclusion

The conducted data-driven analysis gave an overview of airline ticket price behavior for two largest cities in Russia: Moscow and Saint-Petersburg. It was found that price dependence on earliness of ticket purchase differs dramatically for local and international flights, for the period under study (spring-summer of 2015). Prices on tickets to other countries decrease with the increase of time interval before departure to some minimal value. Prices on local tickets in opposite grow. It shows the peculiarity in Russian ticket market and to some extent proves the relevance of empirical researchers devoted to Russian market specifically. Also, there were discovered some interesting patterns in the price volatility. However, determination of reasons requires a deeper analysis to be performed with the use of larger volumes of cleaner data. Two empirical data-driven models were proposed and identified which were aimed to forecast the price according to the earliness of purchase: sigmoidal function approximation was used for domestic flights, and cubic parabola was used for international flights.

Acknowledgements

This work was financially supported by the Russian Foundation for Basic Research, Grant 15-29-07034 "Technology of distributed cloud computing for the simulation of a large city".

References

Barrett, S. D. (2004). How do the demands for airport services differ between full-service carriers and low-cost carriers? *Journal of Air Transport Management, 10*(1), 33-39.

Brander, J. A., & Zhang, A. (1990). Market conduct in the airline industry: an empirical investigation. *The RAND Journal of Economics*, 567-583.

Brons, M. (2002). Price elasticities of demand for passenger air travel: a meta-analysis. *Journal of Air Transport Management, 8*(3), 165-175.

Etzioni, O. (2003). To buy or not to buy: mining airfare data to minimize ticket purchase price. *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, 119-128.

Forbes, S. J. (2008). The effect of air traffic delays on airline prices. *International journal of industrial organization, 26*(5), 1218-1232.

Francis, G., Fidato, A., & Humphreys, I. (2003). Airport–airline interaction: the impact of low-cost carriers on two European airports. *Journal of Air Transport Management, 9*(4), 267-273.

Gillen, D., & Lall, A. (2004). Competitive advantage of low-cost carriers: some implications for airports. *Journal of Air Transport Management, 10*(1), 41-50.

GO TRAVEL UN LIMITED. (2015). *Aviasales API*. Retrieved August 17, 2015, from Aviasales: http://www.aviasales.ru/API

Groves, W., & Gini, M. (2013). An agent for optimizing airline ticket purchasing. *Proceedings of the 2013 international conference on Autonomous agents and multi-agent systems*, 1341-1342.

Hofer, C., Windle, R. J., & Dresner, M. E. (2008). Price premiums and low cost carrier competition. *Transportation Research Part E: Logistics and Transportation Review, 44*(5), 864-882.

Klein, S., & Loebbecke, C. (2000). The transformation of pricing models on the web: examples from the airline industry. *13th International Bled Electronic Commerce Conference*, 19-21.

Lazarev, J. (2013). The welfare effects of intertemporal price discrimination: an empirical analysis of airline pricing in US monopoly markets.

Moscow Domodedovo Airport. (n.d.). *Airport Today - About us*. Retrieved August 21, 2015, from Domodedovo Moscow Airport: http://www.domodedovo.ru/en/main/about/

Njegovan, N. (2006). Elasticities of demand for leisure air travel: A system modelling approach. *Journal of Air Transport Management, 12*(1), 33-39.

Northern Capital Gateway LLC. (n.d.). *About the Airport - Airport Indicators*. Retrieved August 21, 2015, from Pulkovo Airport Saint Petersburg: http://www.pulkovoairport.ru/en/about/performance/

O'Connell, J. F., & Williams, G. (2005). Passengers' perceptions of low cost airlines and full service carriers: A case study involving Ryanair, Aer Lingus, Air Asia and Malaysia Airlines. *Journal of Air Transport Management, 11*(4), 259-272.

Pels, E., & Rietveld, P. (2004). Airline pricing behaviour in the London–Paris market. *Journal of Air Transport Management, 10*(4), 277-281.

Pitfield, D. (1993). Predicting air-transport demand. *Environment and Planning A., 25*(4), 459-466.

Pitfield, D. (2005). A time series analysis of the pricing behaviour of directly competitive'low-cost'airlines. *International Journal of Transport Economics/Rivista internazionale di economia dei trasporti*, 15-39.

Sabre. (2015). *APIs. Get travel information on demand with our REST and SOAP APIs*. Retrieved August 17, 2015, from Sabre Dev Studio: https://developer.sabre.com/docs

Sheremetyevo International Airport. (n.d.). *About us - JSC SIA Key Indicators*. Retrieved August 21, 2015, from International Airport Sheremetyevo: http://svo.aero/en/performance/

Vnukovo International Airport, JSC. (n.d.). *Passenger traffic in 2014*. Retrieved August 21, 2015, from Vnukovo International Airport: http://corp.vnukovo.ru/press/news/passazhiropotok-mezhdunarodnogo-aeroporta-vnukovo-v-2014-godu-vyros-na-13-9/