Svetlana V. Kuzmenkova,
Leading Analyst PJSC BS “Ural-FD”
Aleksey L. Chadov,
Senior Lecturer NRU HSE Perm
2018

Subscribers` Mobility as a Potential Factor of the Cellular Services Consumption

Introduction

According to the AC&M Consulting, from 2000 to 2010 the number of active subscribers has been increased by 7184% and today more than 234 million people in Russia use mobile services. The starting point of this growth is the crisis in 1998 when mobile services companies needed to survive and thus they changed their strategy of operating in the market.

The core elements of the changes were the significant drop in cellular telephones and services prices and development of more complex pricing schemes, aggressive advertising campaigns and setting-up new competition policies. That is why cell operators were able to expand customer base dramatically. Despite the fact that nowadays cellular market is one of the most congested markets, telecommunication providers are still keep to these strategies and interested in maintaining demand and extension of the product range. In order to make these processes work a company should study the preferences of its clients and their behavior. For that reason studying of the demand is still vital for companies.

Cellular services consumption is influenced by a lot of factors. Besides tariff characteristics demography and life style are of the much importance which has an effect on clients` behavior. The last factor represents the great interest for the current work. The lifestyle of subscribers is stays unexplored in the context of cellular demand, and plenty of papers try to throw light on life style characteristics. Miravete (2003) who studies consumption behavior and subscribers` mistakes during tariff plan choice considers some factors which reflect life dynamic, such as the facts of marriage and movement to the other place of living. The paper of Lambreht et al. (2007) introduces a business customer indicator that influences the preference shifter or usage shock of clients. Then, researches that are not related to the cellular market, for instance, studies of human trajectories, sociology and biology, analyze people`s life style by numerical indexes and measures. They suggest different by the nature indices that take into account human movements, and it is called mobility. Among this kind of works Gonzalez et al. (2014) and Hoteit et al. (2014) can be pointed out.

According to these examples, researchers cannot agree on singular measure of people`s lifestyle and life dynamic. By this reason, the major motivation for the current paper is provided by unclear view on lifestyle and mobility measures in the context of subscribers` behavior, and there is a need to bring the fresh breath into the cellular demand consideration and mobility measures.

The main purpose in the current paper is to investigate the effect of subscribers` mobility on their consumption behavior. In order to achieve the goal it is necessary to meet the objectives:

1) Introduce measures of mobility.
2) Analyze subscribers according to their mobility.
3) Build and estimate the choice model.

The understanding of subscribers` behavior and their preferences is an urgent question for cellular companies, especially today when the economic situation in the world changes over time and cellular services market stops to grow. Providers always have to change pricing strategies and marketing policies. Thus, the results of the research are useful for them to develop more sophisticated pricing schemes and correct the existing ones. If the mobility factor does influence the choice of the tariff and consumption, company can create more detailed application form to subscribe services.

The structure of the paper is the following. The first section presents the relevant literature review. In the second section the major aim of the research and hypothesis are described. It is followed by methodology and empirical results, sections three and four respectively, and end up with conclusion in the fifth section.

1. Theoretical background

The related literature is divided into two samples. The first one considers the mobility estimations and the second one represents services consumption modeling approaches.

1.1 Mobility

The concept of mobility can be found in various contexts such as social sciences, economics, biology and neuro science, urban planners, telecommunication and transportation. In recent years, mobile data-based research reaches important conclusions about various aspects of human characteristics, such as human mobility and calling patterns (Gonzalez et al, 2008; Hohwald et al. 2010; Hoteit et al., 2014), virus spreading (Huerta et al, 2002; Wang et al, 2009), social networks, content consumption cartography (Hoteit et al, 2012), urban and transport planning, network design. The availability of detailed movement data allows studying this concept more deeply. Nevertheless there is uncertainty about the measures of mobility and the nature of people`s movements. The first reason for this is that different types of data are used in papers. The GPS data shows more detailed movements and more frequent ones. But the mobile data that is used in the current paper gives information about the location of the subscriber only when he or she makes a connection (a call, sends message, connects to the internet etc.). This does not provide accuracy and fine granularity so that such kind of data is limited and this fact narrows the methods to estimate the mobility.

The literature being discussed relates several papers that introduce techniques to predict user movement between two locations. Wei et al. (2012) and Zheng et al. (2012) created method to identify the most popular travel routes in a city. It is helpful for taxi drivers, for instance, to find customers optimally. The other plenty of papers considers the mobility in terms of trajectories that people walk such as Rhee et al. (2008), Chen et al. (2014), Winter and Yin (2010) and others. The models that estimate the mobility patterns are called mobility models and connected with the concept of random walk and Levy flight. Using GPS traces authors analyze moving trajectories and predict them. These researches use advanced methods to measure mobility, however the data type is more precise.

The other researchers also measure human trajectories and in addition to that they compare people with each other in terms of mobility. The paper of Hoteit et al. (2014) measure the error between real human trajectories and estimated ones and use data type similar to the current paper using different interpolation methods (linear, cubic, nearest interpolations) taking into consideration mobility parameters. Furthermore the error between real and estimated load using the proposed interpolation methods is estimated based on real cellular network activity data of the state of Massachusetts authors found that trajectory estimation methods show different error regimes whether used within or outside the ‘‘territory’’ of the user defined by the radius of gyration. They highlight the fact that people`s trajectories with different mobility can be predicted with different methods. Consequently, this is the evidence for existing difference in people`s behavior with high and low mobility levels.

The index of gyration is also used in (Gonzalez et al., 2008). They analyze the index more precisely. Index of gyration is defined as the deviation of user positions from the corresponding centroid position. The formula is the following:

$$r_g = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\vec{p}_i - \vec{p}_{centroid}\right)^2} \tag{1}$$

$$\vec{p}_{centroid} = \frac{1}{n}\sum_{i=1}^{n}\vec{p}_i\ , \tag{2}$$

where: $\vec{p}_i$ - geographical coordinates of user`s positions;

$\vec{p}_{centroid}$ - center of mass of the user`s positions;

n – number of movements.

The center of mass is presented by average coordinate of user`s positions.

According to the cumulative distribution function of radius of gyration authors distinguish the categories and types of people.

It should be pointed out that in theory centroid position is measured by weighted average in order to overcome the problem of comparability and movement to long distances.

1.2 Demand modeling

The literature on cellular services demand modeling has a great history. Telecommunication works are divided into two samples: the first one considers fixed line networks and the second one – cellular services penetration, competition among providers and clients` behavior. The major interest for the current paper is presented by the demand estimation approaches.

The consumption behavior at the cellular services market is described by several decisions: subscription to a particular provider, choice of the tariff, consumption of services and switching decision. Despite the fact that other stages of decision-making process are equally vital, the current paper analyses second and third ones because of limited to one particular provider`s data.

Several authors build the models that describe multinomial or binary and continuous choices independently for several reasons. Firstly, the data does not contain information about individual consumption. Secondly, this depends on the objectives of the papers. Authors can aim not to model real consumption behavior, but they target to find some effects of consumption, reveal the relationship between services and estimate own- and cross-price elasticities.

Among early papers that study local tariff choice there is Train et al. (1987). The core purpose of the paper is to describe choice of the tariff and its relation to the usage and average duration. The research has a unique feature: it combines two decisions and build fully discrete choice model. The model describes households' interrelated choices of local service option and monthly calling pattern and authors characterize each household as choosing a particular service option and a particular calling portfolio, where a portfolio of calls is defined as a particular number and average duration of calls at each time of day to each distance zone. The estimations are based on Nested-logit specification. The variables considered include fixed monthly payment, per unit fee and geographical time zones. The authors do not include demographic characteristics and subscribers` lifestyle.

The other researches build Logit and Probit models (Albert and Chib, 1993) and Multinomial Logit models (Rossi et al., 1996) that analyze brand choice. However, there is an opinion that continuous decision about the quantity consumed may not be well addressed by discrete models for the reason that it leads to information loss (Niraj et al., 2008) and number of alternatives increases rapidly.

Independent modeling of discrete/continuous choices causes the problem of endogeneity in the case when usage choice is statistically dependent from the tariff choice and vice versa (Kim et al., 2010). The estimates are inconsistent with the presence of endogeneity. That is the reason to analyze two-staged decision process.

The first important implementations of descrete/continuous modeling at the cellular market are made by Hanneman (1984) and Hausman (1999). Subscribers make two decisions: tariff choice and usage choice. The feature of the modeling decision making process in these paper is that the decisions are made simultaneously. This principle is also discussed in Miravete (2002). It must be said that the most significant contribution into the cellular demand theory made Evgenio Miravete. His researches that has been published since 2002 describe basic principles of tariff choice modeling and unusual behavioral subscriber`s principles. The developed models in Miravete (2002,2003,2007) are applied to an experiment that was in two cities of Kentucky in the fall of 1986. The idea of the experiment is that in Bowling Green subscribers were mandatory measured service and standard nonlinear pricing model is estimated.

In his early work (Mirvete, 2002) he studies the uncertainty about future consumption that arise because of the time lag between choice of the tariff and choice of the usage and corresponding usage shocks. After the choice has been made a monopolist is able to screen consumers and define ex-post and ex-ante types of the clients in order to lower the degree of uncertainty.

The other paper (Miravete, 2003) concerns mistakes in the tariff choice and learning mechanisms of subscribers. Miravete claim that the probability of tariff choice mistake increases because clients cannot predict their level of usage precisely before they change a tariff. Furthermore, he argues that to a greater extent the choice is driven by subscribers` expectations of their usage level instead of current consumption. The reason is that clients are able to learn their type and estimate the consumption structure. Consequently, their expectations converge to the real usage in the next period, and thus there are fewer mistakes in the future. Using the data described above Miravete estimates various specifications of the model. The first specification includes tariff characteristics and variables that reflect consumption of services directly as second degree polynomial function. The second specification has additional demographic factors and the third one contains benefit from the usage of two-part tariff measure. Then, the author analyses subscribers` behavior under different tariff regimes. He proves that subscribers who choose flat tariff have high level of consumption, and because of that their choice is rather rational for such kind of clients. Lastly, the research confirms the fact that switching between tariffs is the result of cost-minimization problem solving, and actually subscribers are sensitive even to low changes in their billing.

The next paper written by Lambreht et al. (2007) is the extended previous model that accounts for usage uncertainty and switching behavior between internet providers. They use similar type of subscription service data that include detailed demographic characteristics and tariff regimes. In comparison with the researches above, authors include types of subscribers divided by their workplace and lifestyle and build the model in the context of three-part tariffs. They reveal different consumption and psychological effects. The estimation of the factors shows that tariff choice is driven by prices and allowance (number of free minutes) more than usage quantity. Based on results authors derive an optimal pricing strategy for a company.

The other research that was published in 2008 is written by Tahanory and Toshifumi. It presents the demand estimation of cellular services on the basis of Mixed-logit model. The authors prove high substitution effect between different alternatives of one provider instead of similar alternative of the competitors. Moreover, cross-elasticity effects for additional services such as mobile e-mal, GPRS and SMS are estimated.

The paper of Kim et al. (2010) also considers two services consumption. The main purpose of the research is to evaluate substitution effect between voice and SMS. The econometric approach is based on discrete/continuous choice model in the context of three-part tariffs. Another work of Kim (2012) incorporates switching decision into the process of decision making at the cellular services market. He argues that subscribers decide whether they are willing to stay with the current provider or leave it at every time period with the presence of switching costs. Kim uses Berry-Levinsohn-Pakes model simultaneously with Nested-logit model which allows to account for various random effects. He gives the evidence for the fact that the existence of switching costs prevents subscribers to choose the best alternative. Furthermore, he points out that the expectations of an increase in number of providers at the market will lead the switch decisions to be more advantageous because benefits exceed costs.

Overall, the papers discussed contribute core principles and features of the current modeling and estimations. Firstly, it is assumed to analyze measures and choose the most appropriate one also using methods proposed by papers from the theoretical background. Secondly, the discrete/continuous choice model is estimated in the context of various types of tariffs as the provider offers complex pricing schemes for its customers. Because of that the price elasticities are of interest in the paper. Moreover, the price on SMS is observed, so, cross-elasticities are also to me estimated.

2. Research design

The core purpose of the paper is to identify the effect of subscribers` characteristic that is called mobility. The question under consideration has theoretical implication as it is suggested to find the best measure for this phenomenon. The empirical implication of the paper concerns the ability of the provider to enhance its current tariff plans. This means that the two-staged model and subsequent information about the factors of the tariff plan choice make the company able to predict the probability of choosing an alternative if it changes some pricing parameters and add characteristic into application form that indicate mobility. The consumption behavior on the cellular services market is explored in the context of tree-, four- or two-part tariffs. For this reason estimation of mardinal effects is considered. These results are also useful for company for tariff choice probability and consumption prediction.

The data set is provided by Perm telecommunication company “Rostelecom”. The company has long history at the cellular market in Perm Region being renamed from “UralSvyazInform” and “Utel”. Now the company has merged with large Russian telecommunication company “Tele2” and is continuing to penetrate Russian cellular market. According to AC&M Consulting near 50% of Perm Region subscribers uses Rostelecom services.

The data contains detailed information about individual consumption. For every active subscriber each service that he or she made, exact time and date of the connections, prefix of the conversation partner, packages, costs of every connection are known. The most striking and interesting is that it is know from what exact cell the connections are made. This gives the opportunity to catch the movements of the clients. This information can reflect the lifestyle and life dynamic of the subscribers, thus it is used to measure the factor of mobility.

The empirical results are obtained by three-staged analysis. The first step is to introduce mobility measures and choose the most appropriate ones for the further implication. In order to choose a measure for mobility the criteria have to be specified. Foremost, the measure should provide clear comparability of subscribers in terms of mobility. With the help of the common sense properties of appropriate measures are derived. Secondly, the measures are compared in terms of model quality. The model specification can be compared with the information criteria, pseudo R-squared and value of log likelihood function. The best model with the corresponding measure identifies the best mobility index.

The second step includes model specification and identification. It is necessary to understand the process of decision making. The decision making process of individuals at the cellular services market is described by several stages. Firstly, clients decide what provider they would like to choose.

Then, after they have chosen the provider they subscribe to its services, so they need to choose the tariff that meets their needs. After some period of time clients begin to consume. It has to be bear in mind that subscribers can identify whether their tariff is optimal looking at the billing. Consequently, they can switch the tariff, and the phenomenon of learning is described in Miravete (2003) and switching decisions are presented in Danaher (2013), Lambreht et al. (2007).

The current paper presents two basic decisions of the cellular consumption such as choice of tariff and usage. In order to incorporate the uncertainty about future consumption (Miravete, 2002) the multinomial Heckman model is estimated. The estimation is two-staged. Firstly, the multinomial choice is estimated Multinomial Logit model. In contrast to simple Logit models, the model accounts for the effect of presence of the other alternatives on the choice and because alternative-specific and case-specific regressors are observed, this model is estimated. In contrast, Multinomial Logit accounts only for chosen alternatives and allow using case-specific variables along with alternative-specific. Lastly, the comparison between models is assumed.

After choice estimation, predicted probability for each alternative should be derived. Then, these probability scores enter the second stage equation of the usage represented by Negative Binomial model. The choice of the model is based on the distribution parameters of usage that is out coming calls. Thus, it can found what drives choice of the tariff and usage as well.

The concluding step in the analysis is sampling according to the mobility levels. In the previous research (Kuzmenkova, 2014) different effects of the model parameters on the services consumption have been found. For instance, high mobile people are insensitive to price changes. The groups have been chosen according to the average mobility within tariffs. Consequently, the groups of mobility are connected with the tariffs chosen by clients. The current research uses the same method of sampling.

2.1 Hypothesis

The major interest in the current paper is referred to the mobility factor. This characteristic can reflect the lifestyle and life dynamic of an individual. The data provided contains information about the base station address inside which radius the connection has been made. This gives the opportunity to calculate the distance between cells and total distance at given period of time, frequency of changing cells and the share of connection in each cells. These measures are consolidated into several mobility indexes. The indexes can reflect different sides of the subscribers` mobility, so that the further detailed analysis will show the nature of each measure.

The results obtained in the previous work (Kuzmenkova, 2014) give the evidence for existing of the mobility effect on the services consumption. This research is to test previous results with the extended model. Furthermore, the most appropriate index is to be derived. Consequently, the hypothesis for the mobility is the following:

1) Mobility is characterized mostly by a measure of frequency of changing cells and distance covered. The mobility measure should be aggregate in order to meet all its properties. It should allow for comparing those subscribers, who, for example, change many cells but drive for short distance and change a lot of cells but drive for long distance.
2) High mobile people have different behavior in comparison to medium and low mobile ones. Individuals with dynamic lifestyle tend to react on the changes of tariff schemes and their movement profile less sensitive because they do not have time on it.

The subsequent hypothesis set is divided into two groups: tariff choice and consumption hypothesis. The first stage of the decision making process represents the choice of the tariff. Subscribers can choose among set of tariffs that are provided by the company. The set varies over individuals for the reason that number of available tariffs offered is different for each month. Besides, individuals each month choose whether they stay on the current tariff or switch to the available one at this period of time. All tariffs have different price structure, and when an individual wants to subscribe a service he or she pays attention to the prices, first of all. People tend to choose among not all alternatives, but they group them according to some characteristics and then choose among one group. Commonly, clients outline flat rate tariffs and measured ones, tariffs with monthly fee and without it, tariffs that are internet providers or not and other groups. Moreover, the related papers include pricing scheme and these variables have great influence on the tariff choice. For these reasons the pricing scheme hypotheses are proposed:

3) Mobility has positive effect on the probability of choosing tariffs with monthly fee. This suggestion comes out from papers of Lambreht et al. (2007) and Miravete et al. (2007) who determined the effects of monthly fee payment.
4) Fixed subscription payment and monthly fee decrease the probability of choosing the alternative. This is rather logical, because usually price has negative influence on the demand. Then, according to various authors, the effect of monthly payment is higher than the fixed one.

The second stage is the choice of the consumption. It is suggested that consumption is influenced by the same factors affecting the choice of the tariff and in addition to that, types of the tariffs are incorporated and other services that a client is subscribed to. Moreover, the consumption equation includes Heckman correction term that is additive function of predicted probabilities of each alternative. This term should not to be correlated with residuals of the model in order to the equation be identified. The consumption hypotheses are the following:

5) The consumption of the other services decreases the voice usage as they tend to be substitutes. Kim et al. (2010) and Andersson (2006) estimated SMS service to be substitute for voice. The papers being discussed in the previous section do not contain observations on optional services such as Internet and MMS, that are going to be included into the model.

The hypotheses are tested by the econometric approach. The two-staged decision model is estimated. Moreover, different specifications are compared on the basis of various criteria. Then, some the mobility and prices hypothesis are tested by data analysis and sample estimations.

2.2 Model assumptions

In order to build a model and identify it in the context of the current data, it is necessary to point out some induced theoretical assumptions based on related literature and intuitive assumptions.

1) Mobility does not reflect broad picture of movements. The mobile data is rather limited because subscribers are observed only when they make a connection.
2) The test sample is constructed with the help of filtration of the clients, who move out from the Perm less than 30% times and the duration of services there is less than 2%. For such subscribers these movements and connections can be thrown away without loss of generality.
3) For the proposed models, individuals make choices under the utility maximization assumption. This means that subscribers choose the alternative that brings them maximum utility in comparison with others.
4) Absence of switching costs between tariffs.
5) The paper does not consider switching decisions among provider and within one provider. Although, the absence of these decisions can bias the results, they are left out for the simplicity of analysis.
6) All individuals decide about the tariff on a monthly basis. Firstly, provider changes tariff characteristics also at least not each week. Secondly, the aggregated data smoothes the difference in consumption in the afternoon and in the evening.
7) If an individual switch to another tariff plan during a month, this decision is not taken into account. Instead, the first tariff is treated as the choice this month. This assumption is a consequence of the previous one. Characteristics of the first tariff at given month are suggested to have more influence on the tariff choice in contrast to the last tariff which parameters influence switching decision at greater extent.

8) The choice set varies within individuals at each point in time. It is composed of available tariff plans that are different each month plus the previous tariff if it is not available in the current period but the subscriber can stay on it this time.
9) The marginal prices in the case of complicated tariffs are calculated by the average price for the corresponding service and option. For instance, the price for calls within the provider is differentiated according to the cumulative expenditures at current period of time: if the expenditures less than 200 rubles, the price is 1,2 rubles per minute, if the expenditures lay between 200 and 600 then the price is equal to 0,5 rubles. So, the average price for the network calls is 0,85 rubles. The assumption is driven by unavailability to account for this pricing structure on monthly basis.
10) Subscribers decide only about out coming calls at the second stage. Other services are entered the model in the form of regressors. Actually, clients decide about the consumption of all services simultaneously. Consequently, the second stage has to represent simultaneous equations, but this assumption is relaxed for the simplicity of the estimation process.

Some of the assumptions can be treated as limitations of the research, but this is the opportunity for the further analysis of the subscribers` behavior.

3. Methodology

This section represents methodological approach to the results of the research. Methodology is divided into three aspects: mobility measures, model design and data analysis.

3.1 Mobility measures

Cellular demand literature does not take into consideration the factor of mobility. This is vital subscribers` characteristic for the reason that it describes movements in and out the city, the life dynamic and lifestyle in some sense. From one point of view, it can reflect demographic characteristics, for example, age of clients, which are not observed in the current paper. From the other point of view, it may serve as an additional factor that explains services consumption and clients` behavior. Thus, there is a question raised: how can mobility be measured?

Several approaches for calculating mobility are represented in this paper. The data set contains information about the address of each cell, in radius of which the connections have been made. The cell by definition is a geographic area of base station or cell site signal coverage that takes the form of a circle. The working range of a cell site that is the radius of a signal can vary because of different factors. The information needed for deriving a measure contains the distance between the cells, number of connections in each cell and frequency of changing cells. First easy approach is to summarize the quantity of calls and SMS to find a frequency of being connected during a day for each client or the total distance he covered. Using this approach is not proper because mobility is described by both frequency and distance. Second way is to use different average values and weighted averages that also very approximate the estimation. The other method is to use indexes. The concentration indexes can probably describe the mobility. The more concentrated the calls and SMS are in one or more cells, the less mobile the subscriber is because he does not move to another cell to make a call. Some measures were introduced in the previous paper Kuzmenkova, 2014), thus in the current paper the emphasis in the description is made on new proposed methods.

1) Frequency of changing cells

The measure reflects number of cells that a client has been to and ranges between unity and infinity. However, the minimum value of the measure is disadvantageous because it is incomparable with the indexes starting with zero. For that reason it is sensible to throw away one cell to make this measure a real frequency of changing cells. The formula is given below:

$$M_{it} = |\overline{K}_{it}| - 1, \tag{3}$$

where: $i = 1,..., n$ – individuals set;

$t = 1,..., T$ – time period set;

$\overline{K}_{it}$ – ordered by the date list of the cells of i-subscriber at time period t;

$|\overline{K}_{it}|$ – number of cells for subscriber i at time period t .

The other disadvantage of this measure is that it does not account for the distance covered. If the client moves longer distance for particular time period than the other frequency of changing cells being equal, then this client is more mobile.

2) Total distance

This measure reflects the distance covered during some period of time. The more is the distance, the higher is mobility. The formula is the following:

$$M_{it} = \sum_{k \in \overline{K}_{it}} S_{k,k+1} \quad , \tag{4}$$

where: $k = 1,..., K$ – set of cells;

$\overline{K}_{it}$ – ordered by the date list of the cells of i-subscriber at time period t;

S – distance measured in meters from k-cell to k+1-cell for i-subscriber at time period t.

The measure can be appropriate for clients with equal frequency of changing cells. Otherwise, the usage of total distance itself leads to underestimation or overestimation of people`s mobility as well as the previous measure. Consequently, the aggregate index is proposed.

3) Weighted average distance per day

This measure overcome the problem of two previous ones and account for distance and frequency. The index calculates mobility on daily basis and show the average distance per day at given month. It is presented below:

$$M_{iw} = \frac{\sum_{t=1}^{T} \tilde{S}_t \cdot f_t}{\sum_{t=1}^{T} f_t}, \tag{5}$$

where: $d = 1,..., D$ – day set;

$\tilde{S}$ – total distance for individual I at time period t.

The index is ranged between zero and infinity. Zero mobility means that a subscriber does not move from one sell. Although this turns into indeterminacy in mathematical sense, this is converted into zero in the data.

4) Index of gyration

Related literature introduce mobility index that is called index of gyration, formulas (1)-(2). It is calculated of the distance basis and reflect the standard deviation of user`s current position from the centroid position for the period considered. The major shotcome of the measure is that the center of mass represents simple average instead of weighted average. If subscriber travels a long distance few times, the center of mass is shifted to the corresponding coordinate points. In this context, the mobility increases with an increase of standard deviation of distance.

5) Concentration ratios

In the context of mobility this measure characterizes the distribution of connection among cells. The more concentrated the connections are in several cells, the less is the mobility because a subscriber stayed in one cell instead of moving to another. From one side, the measure reflects the willingness of subscribers to move. From the other side, this gives different information about the mobility in comparison with movement itself. There are various concentration indexes in the literature. The current paper suggests considering several indexes, and then choosing the most appropriate one for subsequent modeling.

The k cell concentration ratio

Summing only over the shares of k most popular cells for subscriber at certain period of time, it takes the form:

$$CR_k = \sum_{i=1}^{k} s_i \ , \tag{6}$$

where: – s is the share of calls in cell k.

This index of concentration emphasizes the k leading cells and neglects other cells in movements of subscriber. There is no rule for the determination of the value of k, so that the number of cells included in the concentration index is a rather arbitrary decision. The index ranges between zero and unity, it approaches zero for an infinite number of equally important cells (given that the k chosen for the calculation of the concentration ratio is comparatively small as compared to the total number of cells) and it equals unity if the cells included in the calculation of the concentration ratio make up the entire movement profile.

If leading cells have large share of connections, then the movements are highly concentrated in these cells, consequently the mobility is low. Contrarily, if connections are distributed uniformly, other things being equal, the concentration is lower and the mobility is higher. An increase in number of cells in subscriber`s movement profile leads to a fall in leading cells share, so that the mobility is increased.

The ignorance of less popular cells makes this measure rather approximate. For instance, if we compare two subscribers, one has ten cells in movement profile with four cells forming 70% of connections. Another has only six cells and leading ones form also 70%. The index in both cases shows equality of concentration, thus the mobility of subscribers is the same. However, mobility of the first subscriber should be higher for the reason he has more cells in movement profile, then he makes more movements than the second subscriber. It can be seen that the index does not take into account number of unpopular cells. That is why it underestimates the mobility.

The advantageous side of this index concerns the popularity of the cells for the provider to see, where subscribers appear more often, and decide about the marketing policy and promotion of new plans at a certain territory.

The Herfindahl-Hirschman Index

The Herfindahl-Hirschman Index (HHI) is the most widely treated summary measure of concentration in the theoretical literature and often serves as a benchmark for the evaluation of other concentration indices. Often called the full-information index because it captures features of the entire distribution of cells sizes, it takes the form:

$$HHI = \sum_{k=1}^{K} s_i^2 \,, \tag{7}$$

It is the sum of the squares of cells sizes measured as shares of voice and SMS. Instead of concentration ratio this index account for all cells in movement profile that makes it more appropriate for measuring mobility. The *HHI* index ranges between 1/*n* and 1, reaching its lowest value, the reciprocal of the number of cells, when all cells in a movement profile are of equal size (concentration), and reaching unity in the case when subscriber makes calls and send SMS only from one cell. This index can be represented through the connection between number of cells and variance of shares of connections to the network:

$$HHI = (1/K) + K\sigma^2 , \tag{8}$$

The connection with variance provides the index to be sensitive to the changes in shares of calls and SMS in cells. If asymmetry is large, the value of index is high. But in the context of mobility an increase in the dispersion of the connections while number of cells stays unchanged does not mean an increase in movements and thus mobility. This index has significant implication for the identifying the popular cells that can be important to the company`s marketing policy. The effect of extension of profile is positive: the more cells the subscriber has, the less is concentration as connections are distributed to a larger number of cells, the more is mobility.

Herfindahl-Hirshman index overcome the limitations of the concentration ratio. It accounts for all of the cells and, moreover, the exact effect of the shares variance and number of cell can be extracted. Nevertheless, the great shortcome of concentration indexes is that they do not consider all movements. They reflect only aggregate number of connections in cells and cells the subscriber use during a period. Consequently, it cannot be observed how many times a client moves.

There are various indexes of concentration but still it can be said that Herfindahl-Hirshman index is the most appropriate one because it incorporates the inequality of connections distribution among cells and is connected with the dispersion of the shares. In contrast with this index, the concentration ratio underestimate unpopular cells and comprehensive industrial index overestimate the most popular one. The last conclusion remark is connected with the limitation of such indexes. Concentration index takes into account unique cell, so the set of the cells no longer represents the ordered list of subscribers` movements. Consequently, the movements itself are not caught and this measure feature different factor of subscribers behavior.

3.2 Modeling approach

The current paper models consumption behavior by discrete/continuous choice model. This class of models is developed to combine the choice of discrete alternatives and choice of continuous factor and correct for nonrandom sampling. They are called sample selection models and has been introduces by Heckman (1974, 1979). However the classic Heckman model accounts for a binary choice that is two alternatives are suggested. For the reason that the current research considers more than two alternatives, the multivariate extension of Heckman model is presented in Das et al. (2003).

The intuition behind sample selection models is that two decisions can be dependent, so the errors of tariff choice and choice of the usage are correlated. The proposed model corrects for the sample selection using the vector of probabilities of each outcome (Cameron and Trivendi, 2009).

First stage – choice of the tariff

The first stage of the model is presented by unordered multinomial choice model. The categories are presented by a set of 54 tariffs. This class of models is considered in order to predict the probability of each outcome which is of interest of the current provider. The core principle of choosing an alternative is the utility maximization. The subscriber will choose the tariff that brings him or her higher utility than any other alternatives.

Let denote y_i as the outcome for i individual of J alternatives. Then, set $y_i = j$ if the outcome is the *j*th alternative, $j = 1, 2, \ldots, J$. For individual *i* and alternative *j* the utility function U_{ij} is constructed as an additive function of regressors and error term ε_{ij}. Regressors are divided into alternative-specific, x_{ij}, which vary within alternatives, and case-specific, z_i, which vary across individuals and are invariant across alternatives.

$$U_{ij} = x'_{ij}\beta + z'_i\gamma_j + \varepsilon_{ij}, \tag{12}$$

where: β and γ_i are the parameters to be estimated.

As it has been mentioned above, the outcome $y_i = j$ if j provides the highest utility. Then, the probability of choosing an alternative can be written:

$$\Pr(y_i = j) = \Pr(U_{ij} \geq U_{ik}) = \Pr(U_{ik} - U_{ij} \leq 0) = \Pr(\varepsilon_{ik} - \varepsilon_{ij} \leq x'_{ik}\beta + z'_i\gamma_k - x'_{ij}\beta + z'_i\gamma_j), \forall k \tag{13}$$

In order for the probability to be estimated the assumption about idiosyncratic term should be made. Commonly, ε_{ij} is treated as independent and identically distributed random term with logarithmic distribution of type I extreme value (Ben-Akiva, Lerman, 1985), (Nevo, 2001), (Cameron and Trivedi, 2009).

The estimation results cannot be interpreted directly. This means that the coefficient with the positive sign do not lead to an increase in the probability of choosing an alternative. Instead of it, marginal effects should be calculated that reflect the change in probability caused by the change in variables. The formula is given below:

$$ME_{ijk} = \frac{\partial \Pr(y_i = j)}{\partial x_{ik}}, k = 1, \ldots, K, \tag{14}$$

$$ME_{ijp} = \frac{\partial \Pr(y_i = j)}{\partial z_{ijp}}, p = 1, \ldots, P, \tag{15}$$

where *k* is the set of alternative-specific regressors and *p* is the set of case-specific regressors.

Different multinomial models exist due to different assumptions about the joint distribution of error terms, $\varepsilon_{i1}, \ldots, \varepsilon_{iJ}$, with different specifications of probability function. The current paper considers Multinomial Logit. The choice of these models is driven by the ability to include all types of variables and by ability of each alternative prediction which is needed for the next step to be identified.

The studies that explain the choice only by case-specific variables suggest estimating Multinomial Logit. Despite this fact, alternative-specific regressors can be included. This is the simplest model and the parameters are easy to interpret. The disadvantage of the model is that it suggests the independence among alternatives and does not take into consideration the actual choice set. Consequently, Multinomial Logit model requires the data to be in wide form, that is one chosen alternative is observed for an individual.

Predicted probability takes the form:

$$p_{ij} = \frac{\exp(x'_{ij}\beta + z'_i\gamma_j)}{\sum_{l=1}^{J}\exp(x'_{il}\beta + z'_i\gamma_l)} \quad . \tag{16}$$

The model ensures that the probability of choosing an alternative varies between zero and unity and the sum of probabilities is equal to 1. In the ML model there is a base alternative that provides the identification. The coefficient of this alternative is set to zero and the subsequent interpretation is made in comparison this outcome.

Second stage – choice of the consumption

The continuous choice of usage is presented by nonlinear class of models. According to the distributions parameters of the dependent variable (duration of calls) Negative Binomial model is chosen. The model is based on the count nature of the variable and overdispersion in the data: dispersion exceeds average value by nearly two times. This can relates to the unobserved heterogeneity of the clients.

If y denotes the dependent variable and v is an overdispersion parameter that has Gamma(1, α) distribution where α is the variance parameter of it and μ is the Poisson parameter, then the distribution of y is the Poisson-Gamma mixture that is Negative Binomial distribution, NB(μ, α). The correspondent probability mass function is:

$$\Pr(Y = y \mid \mu, \alpha) = \frac{\Gamma(\alpha^{-1} + y)}{\Gamma(\alpha^{-1})\Gamma(y+1)} \left(\frac{\alpha^{-1}}{\alpha^{-1} + \mu}\right)^{\alpha^{-1}} \left(\frac{\mu}{\alpha^{-1} + \mu}\right)^{y}, \tag{17}$$

$$\mu = \exp(x'\beta + z'\gamma)$$

Furthermore, the equation should incorporate the predicted probabilities from the first stage in order to correct for sample selection and address the dependency between two choices. In order for the results to be identified and to be consistent, the first equation should have excluded instruments for each alternative from the second one for the errors to be uncorrelated (Das et al., 2009).

The models are estimated by Maximum Likelihood method.

Last but not the least, the estimation parameters should be mentioned. The models include the following variables (table 1):

The pricing scheme of the tariffs and mobility factor are described in the previous section. Besides these characteristics the paper introduces number of months and number of tariffs for each individual. Number of months reflects the constant characteristic that does not vary across individuals. Number of tariffs reflects the number of tariffs each month. Then, it is interesting to examine the difference in behavior of subscribers, who consume MMS, Internet and are subscribed to packages, although the last variable is a tariff option. The tariff dummies are needed to be introduced into the first-stage equation to obtain the proper prediction, so they are the excluded variables.

Table 1

Alternative-specific and case-specific regressors

Alternative-specific variables	Case-specific variables
Fixed subscription payment	Mobility
Monthly fee	Number of months
Free minutes	Number of tariffs
Marginal fees for calls	GPRS dummy
Marginal fees for SMS	MMS dummy
Tariff type dummies	Packages dummy

It is suggested that the mobility influence both discrete and continuous decisions. More mobile people tend to choose tariffs with fixed payments and consume more.

3.3 The Data Analysis

The next point is the data analysis. The current paper uses data set provided by Perm Telecommunication Company “Rostelecom” that is the largest cellular services provider in Perm Region. Data set contains detailed daily information about individual consumption of all active subscribers during 11 month period from January 2012 to November 2012. The following information is included: tariff regimes, exact date and time of each connection to the network, type of a service, packages for a tariff and exact cell from which connection has been made.

For the reason that the address of the cell station is known the locations of clients can be catch over the time. That gives the opportunity to sample subscribers according to their movements and extract individuals that move in and out the city and others who locates only in the city. The analysis of the test sample of 10000 subscribers shows that 0,021% of clients move only in the city. The current paper considers those subscribers whose movements from the city account for 30% out of total movements and average duration of connections in these cells is lower than 2% of total duration. Thus, the sample is 4550 subscribers. The criteria for sampling is derived though location analysis, particularly histograms. Because outside locations contribute to the small part of total number of locations and duration, such kinds of movements can be dropped out of the consideration without loss of generality. Another reason for it is that for those subscribers who drive rarely outside the city mobility can be miscalculated in terms of proposed methods.

Hence, the initial data set is structured as panel data which is monthly aggregated, but in the research it is treated as cross-section, and it accounts for 47063 in the wide form.

The analysis of locations and base stations takes the significant part of the data analysis in the current paper. The total number of cells or base stations is equal to 3169. The data set captures all connections of the subscribers even if they are not in the Perm Region. For this reason number of cells in Perm Region accounts for 734 and in Perm 272.

The next point is to describe provider`s pricing strategy. Throughout the period of 11 months company offered 65 tariffs. The number of tariffs varies each month as the provider closes tariff plans that are not profitable and open new ones. The overall trend of number of tariffs has decreasing nature. It is reasonable to consider only available tariffs each month (from 21 to 11) and current subscribers` tariffs even if they are unavailable or closed at the time period under analysis, because a client can choose to stay on this closed tariff or switch to available one.

Provider offers wide range of tariff plans that consists of various types. There are not only typical tariff schemes such as two-part, three-part and flat ones. Firm also developed more complex regimes that include four parts and are differentiated by volume consumed or cumulative expenditures schemes. The type of the tariff is set according to the combinations of included charges. Thus, nine types of tariffs are indicated.

Appendix 1 presents the tariffs distribution. It can be seen that 12 tariffs have more than 100 users and they contribute almost 90% of provider`s revenue. In addition to that, there are unpopular tariffs with the consumption percentage less than 0.2%. It is reasonable to drop these clients because the estimations for such tariffs will not be consistent. Consequently, the total number accounts for 36 alternatives.

The descriptive statistics is presented in the table 2.

The most interest is presented by mobility measures. As it can be seen, the distribution parameters vary within indices. For total distance, average distance per day and gyration standard deviation exceeds the mean, but for gyration the difference is larger – the increase is two times. By contrast, frequency and Herfindahl-Hirshman index have large mean. According to these characteristics, measures present different features of mobility.

Table 2

Descriptive statistics for count, continuous and dummy variables

Variable	Observations	Mean	Std, Dev,	Min	Max
Out coming calls	46861	21251,45	27622,55	0	655066
Distance	46861	326,972	388,95	0	6073,653
Average distance	46640	21,842	26,275	0	713,041
Gyration	46640	1365,734	2157,06	0	7784,565
Frequency	46861	43,749	34,756	0	243
HH	46861	2823,223	2032,187	,0119182	10000
Fixed subscription fee	46861	297,045	604,142	0	2000
Monthly payment	46861	27,876	75,241	0	1350
Free minutes	46861	16,032	155,842	0	3000
pc1	46861	0,7444	0,493	0	3
ps1	46861	1,413	0,191	,825	1,5
Number of tariffs	46861	1,244	0,69	1	5
Number of months	46861	10,673	1,236	1	11

Dummy variables (if outcome=1)

Variable	Frequency	Percent
Monthly fee	2228	48,24
Package	1078	23,34
MMS	534	11,56
GPRS	2408	52,13

On average, subscribers move for 44 cells by 323 kilometers during a month. If the distance and frequency are a combined measure, then clients drive for 22 kilometers a day. The maximum value of average distance and distance a month is 713 and 6074 km respectively.

The next position is presented by tariff characteristics. It can be seen, that tariffs has also strong variation. For instance, the standard deviation of fixed payment exceeds the mean by more than two times. The same situation is about monthly fee. The variation is 75 rubles and the mean is 27 rubles. Then, clients pay 0,74 rubles per minute of calls and 1,4 rubles per SMS, on average.

The other factors point out that the number of tariffs and number of months the client is observed varies. So that client can have 5 tariffs during a month.

From the table for dummy variables it can be seen that in 2012 almost half of subscribers consume Internet. The less popular are packages and MMS services (23,34 and 11,56 responsively). In order to account for the effect of subscribers` types who choose the tariffs with monthly fee, the correspondent dummy is included. It shows that half of the clients subscribes to tariffs with monthly fee.

Lastly, the dependent variable of the second stage equation needs to be considered. The average user consumes more than 20 thousand minutes per month and the maximum value exceeds 400 thousand. Moreover, it can be seen that the minimum minutes consumed is equal to zero. The reason is that some out coming calls can be made on the individual`s account for whom the call is made. The other reason comes out from the tariff scheme when calls for the same tariff are free of charge.

The vital point in the data analysis is the correlation of factors (Appendix 2). The estimation can be biased if high correlated variables are included into the model. Eventually, some mobility indexes are highly correlated such as frequency and distance, HH index and distance. This means that including these factors together is not proper. Then, it has to be pointed out that prices are also correlated. Then, the dependent variable of calls` duration is highly correlated with incoming calls and all mobility indexes, but the level of correlation is not critical. Thus, the variables can be included into the model.

4. Results

The first estimation results are presented in the Table 3 that illustrates five specifications of Multinomial Logit according to the mobility indexes for 2 tariffs that take the second and the third place among subscribers. The first specification refers to the distance, second – to the frequency, third reflect the effect of weighted average distance, forth – radius of gyration and fifth is the Concentration Index. Results are robust to heteroscedasticity and autocorrelation biases.

The base alternative is the first-popular tariff for subscriber. Firstly, the choice is based on the suggestion that if the provider wants to launch new tariff plan and close existing one he, would to know the probability of choosing the alternative relative to the most popular one. So that, the comparison is logically made to the most attractive tariff from the subscribers` point of view.

Mobility measure identification

The results are shown only for four tariffs out of 24 because they are the most popular and it is impossible to combine five specifications for every tariff in one table. Moreover, these estimations are presented to choose the most appropriate measure for mobility according to the properties of the models.

According to the estimations (table 3) only HH index is significant for four tariffs. The distance is significant in the last specification for tariff with monthly fee. In the full table of the results that contain all tariffs mobility measures varies in their significance. In order to understand whether there is a real effect of mobility on the tariff choice the joint significance Wald test should be implemented. For all variables the test shows 1%-level of significance within all alternatives.

The number of observations and variables is equal, the variables are significant, and thus this gives the opportunity to compare the models. According to all criteria the first specification with distance is better. Pseudo R squared (52,2%) in first specification is still higher than in other specifications. Information criteria are also lower in distance specification.

Robust check and sampling

Before interpreting tariff choice results, the robust check has to be conducted. In order to do that the sample is cut down with the reference to number of tariffs. The checking sample consists of top 12 popular tariffs. 5 out of 12 tariffs have monthly payment. Thus, the number of observations is 40821 that is 87,11% of the primary sample. This means that dropping 15 tariffs will probably not bias results. The estimations for four tariffs under consideration are presented in the Appendix 3. The variables do not change the significance and signs. Consequently, the first results are robust to changes in alternatives sample. The best model out of these is the second one, according to information criteria of Akaike and Schwartz.

Table 3

Estimation results for five mobility indexes specifications with 4 out of 25 tariffs presented

Variables	Tariff 37					Tariff 20				
	(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)
Distance	-0,0000215					0,0000645				
Average dist		-0,000570					0,00106			
Frequency			-0,000663					0,000207		
Gyration				-0,000018					-0,0000111	
HH					-0,0000240*					-0,0000267**
Fixed fee	0,000535***	0,000493***	0,000440***	0,000491***	0,000508***	0,00276***	0,00273***	0,00269***	0,00223***	0,00274***
Monthly fee	-0,000997	-0,000522	0,0000404	-0,00056	-0,000805	-0,00639***	-0,00502***	-0,00329***	-0,00523***	-0,00576***
Free minutes	0,0168**	0,0114*	0,00443	0,0107*	0,0140**	0,0511***	0,0378***	0,0222***	0,0470***	0,0455***
pc1	0,297***	0,308***	0,332***	0,314***	0,314***	-1,027***	-0,943***	-0,859***	-1,322***	-0,989***
ps1	-6,465***	-5,892***	-5,198***	-5,935***	-6,219***	-3,089***	-2,770***	-2,355***	3,345***	-2,934***
Number of tariffs	-0,201	-0,180	-0,153	-0,183	-0,201	-0,223	-0,164	-0,0869	-0,0120	-0,212
Number of months	0,0787***	0,0794***	0,0768**	0,0782***	0,0754***	0,0501***	0,0487***	0,0482**	-0,0277	0,0471***
Package	0,713***	0,679***	0,623***	0,663***	0,674***	1,152***	1,131***	1,138***	0,463***	1,143***
MMS	-0,0603	-0,0531	-0,0494	-0,0539	-0,0551	-0,0667	-0,0642	-0,0551	-0,00309	-0,0572
GPRS	-0,203***	-0,201***	-0,209***	-0,205***	-0,219***	-0,0870*	-0,0947*	-0,0906*	0,121**	-0,0970*
Constant	8,156***	7,319***	6,313***	7,402***	7,906***	3,877***	3,280***	2,571***	3,345***	3,745***
Observations	46861	46861	46640	46640	46861	46861	46861	46640	46640	46861
Pceudo R-sq	0,524	0,499	0,464	0,492	0,510	0,524	0,499	0,464	0,492	0,510
AIC	127342,5	133889,0	142629,2	135185,2	130965,4	127342,5	133889,0	142629,2	135185,2	130965,4
BIC	130861,9	137312,2	146164,3	138475,3	134301,0	130861,9	137312,2	146164,3	138475,3	134301,0

Variables	Tariff 42					Tariff 43				
	(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)
Distance	0,0000691					-0,000288***				
Average dist		0,00224*					-0,00436***			
Frequency			0,000583					-0,00413*		
Gyration				0,0000574***					-0,0000152	
HH					-0,0000318*					-0,000000412
Fixed fee	0,000664***	0,000598***	0,000536***	0,000614***	0,000633***	-0,000250***	-0,000205***	-0,000168***	-0,000224***	-0,000241***
Monthly fee	-0,0124***	-0,0102***	-0,00727***	-0,0101***	-0,0114***	-0,0107***	-0,00876***	-0,00658***	-0,00907***	-0,00992***
Free minutes	0,00454	0,000203	-0,00391	0,000843	0,00322	0,0626***	0,0469***	0,0291***	0,0721***	0,0570***
pc1*	0,730***	0,733***	0,682***	0,701***	0,722***	-1,810***	-1,729***	-1,563***	-1,673***	-1,725***
ps1	-0,467***	-0,364**	-0,186	-0,331*	-0,404**	-2,203***	-1,887***	-1,562***	-1,969***	-2,100***
Number of tariffs	-0,00110	0,0214	0,0453	0,0240	-0,00363	-0,231	-0,204	-0,156	-0,196	-0,210
Number of months	0,0915***	0,0896***	0,0926***	0,0882***	0,0886***	0,0654***	0,0690***	0,0608**	0,0604**	0,0614**
Package	-0,341***	-0,325***	-0,211***	-0,246***	-0,324***	2,655***	2,672***	2,587***	2,557***	2,583***
MMS	-0,0548	-0,0617	-0,0493	-0,0523	-0,0459	-0,158	-0,138	-0,137	-0,162	-0,169
GPRS	0,267***	0,247***	0,259***	0,247***	0,254***	-0,0986	-0,0744	-0,0932	-0,113*	-0,117*
Constant	-1,069**	-1,323***	-1,587***	-1,337***	-1,018**	3,744***	3,183***	2,498***	3,237***	3,462***
Observations	46861	46861	46640	46640	46861	46861	46861	46640	46640	46861
Pceudo R-sq	0,524	0,499	0,464	0,492	0,510	0,524	0,499	0,464	0,492	0,510
AIC	127342,5	133889,0	142629,2	135185,2	130965,4	127342,5	133889,0	142629,2	135185,2	130965,4
BIC	130861,9	137312,2	146164,3	138475,3	134301,0	130861,9	137312,2	146164,3	138475,3	134301,0

*-p<0,1, *- p<0,05, ***-p<0,01 level significance

*pc1 and ps1 denote the marginal prices for calls and SMS responsively.

The table 4 below presents the marginal effects on the first and second stage estimations. Marginal effects of Multinomial Logit aggregate the coefficients and account effects for each alternative separately. So, the results of the first stage are presented for the base alternatives. The core estimation results are reflected by the sampling analysis of observation according to the mobility levels by tariffs. Table 1 of Appendix 3 presents the summary statistics for consumption by tariffs. Tariffs then can be aggregated into three groups: high, middle and low mobility. Table below contains full and sample specifications of marginal effects.

Within small number of tariffs pricing scheme predict the tariffs perfectly, thus they were excluded from the tariff choice specifications. The significance of variables does not change across all specifications.

According to the full sample results of tariff choice, it is seen that distance does not have the effect on it. The significance is opposite to the second stage equation. The prices preserve the expected signs, however monthly fee has positive sign. This is consistent with Lambreht et al. (2007) because of the presence of psychological effects. Previously, it was found that distance is the best measure for mobility. Thus, **Hypothesis 1 is rejected**, because the measure should reflect the frequency of movements as well. The choice of the distance as best mobility measure can be overestimated by the model, because it has a lot of correlated variables, especially pricing scheme.

More importantly, the results above suggest that subscribers are different in behavior thus the **Hypothesis 2 is not rejected**. First of all, the choice of tariff among groups of subscribers is influenced by distance in different ways, although the **Hypothesis 3 is rejected** for the reason that variable has different signs in different equations.

In particular, medium mobile people have positive influence of distance covered on the probability tariff choice in comparison to the most popular one. However, the value of the coefficient is rather low. High mobile people are insensitive to their life dynamic in contrast to low mobile. The increase in average distance per month by 1 km leads to the decrease of probability of choosing an alternative by 0,7 percentage points for this kind of individuals. In addition to that, the signs of parameters for low type are negative except for number of months and package. The differences in estimations are explained by the difference in clients` behavior. Moreover, different base alternatives can provide the variation in estimations.

The second stage equation represents the significance of distance in all specifications. The variation in the sign also can be explained by the difference in base outcome in first equation for medium mobile clients. Then, the prices influence the consumption for groups in different way. The increase of fixed fee and monthly payment for high mobile people causes the consumption growth. For the marginal price the effect is the opposite.

Table 6

Estimation results of mobility specifications

Variables	Choice of tariff				Choice of usage			
	MNL	MNL high	MNL medium	MNL low	NB	NB high	NB middle	NB low
Distance	$-2,4 \cdot 10^{-7}$	$-9,5 \cdot 10^{-6}$	0,00006**	-0,007***	11,77***	-4,048**	9,404***	-2,032*
Fixed fee	$-6,4 \cdot 10^{-6}$ ***				-0,836	63,892***	-69,225***	-2,25***
Monthly fee	-0,0002***				351,747***	307,393***	-39,919***	
pc1*	-0,0137***				-34921,1***	-23314***	-24855,1***	-40846,2***
ps1	0,0149***				8157,79***			
Number of tariffs	0,0005***	0,038	0,222	-0,0124***	372,05*	-10209,9***	-1427,81***	4234,18***
Number of months	-0,0007***	-0,004**	-0,027***	0,034***	1628,37***	4842,96***	923,273***	1295,56***
Package	-0,002***	-0,104***	-0,242***	0,479***	36861,8***	131522***	192515,5***	88056,8***
MMS	0,0004**	0,034***	0,071***	-0,099***	5966,79***	-4326,71***	5297,335***	3125,12***
GPRS	0,0003***	0,037***	0,034***	-0,133***	2041,41***	-12877,6***	-1636,34***	-8057,85***
p1					-1935,34**			-37043,1***
p51					-34823,1***			
p41					17703,3***		-56157,6***	
p20					-3319,45***			
p34					-1113,49***			
p37						-270485***		
p42					22924,6***	-203867***		
p43					-14947,8			-188162,7*
p52							-77505,7***	
p53								-158145,8
p47							-31170,2***	
p40						-108179,5*		
Observations	40821	17305	14711	8805	40821	17305	8805	14711
Pseudo R-squared	0,99	0,023	0,326	0,317	0,012	0,006	0,014	0,008
AIC	242,0	45992,1	26504,0	16280,5	874008,6	355620,7	188368,1	328935,0
BIC	1284,7	46155,1	26663,5	16429,2	874172,3	355729,3	188467,2	329033,8

*-p<0,1, *- p<0,05, ***-p<0,01 level significance
*pc1 and ps1 denote the marginal prices for calls and SMS responsively. P1, p51 etc. denote correction term for sample selection. MNL – Multinomial Logit, NB – Negative binomial.

High mobile people depend on marginal price changes most of all, and the sign is negative meaning the fall in consumption. The other feature of high mobility is the negative sign near MMS dummy, so these people are tend to consume less because of an increase in average MMS consumption.

Overall result for choice with the respect to hypotheses is about price influence. **Hypothesis 4** about the effect of tariff regimes **is not rejected**. Fixed payment, monthly fee and marginal price for calls decrease the probability of choosing an alternative and the last one has the greatest influence (1,37 percentage points). In comparison with tariff choice, the influence of monthly payment is positive for two specifications of calls` duration consumption that also can be explained by psychological effects. This suggestion also proves the Hypothesis 2.

The last factors that have to be pointed out are the dummy variables for other services. The presence of packages lowers probability of choosing tariff and raises usage. The results for MMS and Internet vary. Mostly, MMS has positive influence and Internet dummy has positive effect on tariff choice probability and negative effect of consumption in almost all cases. Thus, the **Hypothesis 5 is rejected** for the reason that these services can be either substitutes or complements among different groups of people. For instance, GPRS is treated by all types as substitute. The significance and the sign of the variable can be explained by the fact that in 2012 “Rostelecom” started to launch mobile Internet, and the new service became popular quickly and started to attract more consumers.

Conclusion

The current Russian cellular market is characterized by oligopolistic structure that causes severe competition among companies. In order to attract new customers and hold the existing ones they need to revise pricing strategy over time and create more complex pricing schemes to meet preferences of the clients. Thus, the studying of subscribers` preferences is vital for the telecommunication providers today. According to the researches in this sphere, cellular demand is influenced by a lot of factors. Besides tariff characteristics demography and clients` lifestyle is of the much importance.

This paper has given an account of the reasons to consider the mobility of subscribers in the context of the cellular demand. The study set out to determine the effect of clients` mobility on the services consumption. In order to achieve the goal several steps were undertaken.

Firstly, the paper discussed the major literature related to the mobility estimations and demand modeling on the basis of which hypotheses and modeling approach were presented. Secondly, the project introduced several mobility measures, and each of them reflects the unique feature of mobility. It was conducted that the measure should be an aggregate that reflect distance as well as frequency of changing cells.

Third step was the analysis of the data set provided by Perm Telecommunication Company "Rostelecom" and contained detailed individual information about consumption. The most striking observation was about the exact base station address from radius of which the connections have been made by clients. The paper provides statistical and econometrical methods for data analysis.

The econometric approach was based on the two-staged demand estimations with multinomial Heckman model. The choice of the model resulted from the decision-making process at the cellular services market. The first stage refers to the choice of the tariff. Then, after some period of time a client makes the usage choice conditional on the previous decision. For the reason that the company offers great number of tariffs, the choice of the tariff was described be Multinomial Logit model. The second stage equation presents the count regression model with the duration of calls as the dependent variable, particularly Negative Binomial model.

Returning to the hypothesis posed at the beginning of the study, it is now possible to say that the core purpose was achieved. The paper has found that generally mobility does have the influence on the consumption behavior. The first regression analysis revealed that the distance is the best measure for the mobility, and the further considerations were made according to this measure.

The second step was introduced to show the difference in clients` behavior with different mobility types. One of the most significant findings to emerge from this study is that there is significant difference in effects of model factors on both stages for high, medium, and low mobile people. For instance, high mobile clients are insensitive to change in their mobility in respect to the tariff choice. Furthermore, such people have positive influence of monthly payment on the usage choice in contrast to medium and low mobile individuals.

The evidence from this study suggests that the information about clients` lifestyle can be used for prediction analysis. Particularly, providers can use the information about the lifestyle of subscribers and life dynamic in order to predict the probability of choosing a tariff and the changes in company`s profits that can be obtained with the prediction of the minutes consumption.

The present study, however, makes several noteworthy contributions to existing knowledge of mobility factor and demand cellular theory. Firstly, the paper derives appropriate indexes for mobility measurement using mobile GPS data. Secondly, this is the first time that movements have been used to explore consumption behavior. Thirdly, this research will serve as a base for future studies to predict the subscribers` consumption.

Nevertheless, a number of important limitations should be considered. First, the nature of the mobile data does not give the broad picture of movements. The location is known only when subscribers make connections. That is why the life style of clients is not caught clearly. Secondly, the fact of moving out from the city is not considered. On the one hand, this will bias the mobility estimations because measures are sensitive to rare movements for long distances. On the other hand, the behavior of subscribers is described in limited way.

Thirdly, the paper does not take into account demographic characteristic of subscribers because they are unavailable with the existing data set. Moreover, this information is provided to company in sophisticated way, and clients can give incorrect information about themselves. The next point is connected with the prices calculation.

As in has been mentioned in the Research design section, marginal prices present the average numbers for several tariffs for the reason of monthly aggregated data.

For the reason that the test sample contains 10000 subscribers selected randomly, the results can be different if the number of observations will be increased.

The major fact that is not addressed in this study is that decision-making process in not constrained by two decisions. Switching decision is of much importance, because in order for provider to understand the mechanism by which clients choose the tariff, and to be able to predict the tariff choice, it should understand why clients start to prefer the other alternative more or why they leave the provider (Danaher, 2013).

This research has thrown up many questions in need of further investigation. The overcoming of these limitations might be motivation for the future work.

References

Papers**:**

1. Albert, J., H. and Chib, S. (1993), "Bayesian analysis of binary and polychotomous response data" *J, Amer, Statist, Assoc*, 88(422) 669–679.
2. Chen, X., Pang, J. and Xue, R. (2014). "Constructing and comparing user mobility profiles". *ACM Trans*. Web 8, 4, Article 21.
3. Gonzalez, M., Hidalgo, C. A., and Barabasi, A.l. (2008), "Understanding individual human mobility patterns" Nature 458, 238.
4. Grzybowski, L. and Pereira, P. (2008), "The complementarity between calls and messages in mobile telephony", *Inform. Econom. Policy,* 20(3) 279–287.
5. Hanemann, W. M. (1984), "Discrete/continuous models of consumer demand" Econometrica 52 541–561.
6. Hausman, J. (1999), "Cellular telephone, new products and the CPI" J. Bus. Econom. Statist. 17 188–194.
7. Heckman, J. J. (1974), "Shadow Prices, Market Wages, and Labor Supply", Econometrica, 42, 679–693.
8. Heckman, J. J. (1979), "Sample Selection Bias as a Specification Error", Econometrica, 47, 153–161.
9. Hoteit, S., Secci, S., Sobolevsky, S., Ratti, C. and Pujolle, G. (2014) "Estimating human trajectories and hotspots through mobile phone data", *Computer Networks* 64, 296–307.
10. Huerta, R. and Tsimring, L. (2002), "Contact tracing and epidemics control in social networks" Phys. Rev. E 66.
11. Kim, Y., Telang, R., Vogt, B. W. and Krishnan, R. (2010), "An empirical analysis of mobile voice service and SMS: a structural model", *Management science,* Vol. 56, No. 2, pp. 234–252.
12. Lambreht, A., Seim, K. and Skiera, B. (2007), "Does uncertainty matter? Consumer behavior under three-part tariffs", *Marketing science*, 26 (5), pp. 698-710.
13. Miravete, E. J. (2002), "Estimating demand for local telephone service with asymmetric information and optional calling plans", *Rev. Econom. Stud*. 69, 943–971.
14. Miravete, E. J. (2003), "Choosing the wrong calling plan? Ignorance and learning", *American Economic Review*, 93, 297–310.
15. Nevo, A. (2001) "Measuring Market Power in the Ready-to-Eat Cereal Industry" *Econometrica*. Vol. 69, pp.307-342

16. Niraj, R., Padmanabhan, P. and Seetharaman, B. (2008), “A crosscategory model of households’ incidence and quantity decisions” *Marketing Sci.* 27(2) 225–235.
17. Rossi, P. E., McCulloch, R. E. and Allenby, G. M. (1996), “The value of purchase history data in target marketing”. *Marketing Sci.* 15(4) 321–340.
18. Takanori I. and Toshifumi K. (2008), “Discrete choice model analysis of mobile telephone service demand in Japan” *Empir Econ.* Vol. 36, pp.65-80.
19. Train, K. E., Mcfadden, D. L. And Ben-Akiva, M. (1987), “The Demand for Local Telephone Service: A Fully Discrete Model of Residential Calling Patterns and Service Choices”, *Rand Journal of Economics*, 18, 109-123.
20. Wang, P., Gonzalez, M.C. and Hidalgo, C.A. (2009), “Barabasi, Understanding the spreading patterns of mobile phone viruses” *Science* 324 1071–1076.
21. Winter, S. and Yin, Z.C. (2010), “Directed movements in probabilistic time geography”, *Int. J. Geograph. Inform*. Sci. 24, 1349–1365.

Books:

22. Ben-Akiva M. and Lerman S. R. (1985), *Discrete choice analysis*, *The MIT Press*, London.
23. Cameron A.C. and Trivedi P.K. (2009), *Microecomonetrics using Stata*, A Stata Press Publication, Stata Corp LP, Texas.

Conferences:

24. Hoteit, S., Secci, S., Ziemlicki, C., Smoreda, Z., Ratti, C. and Pujolle, G. (2012), “Content consumption cartography of the paris urban region using cellular probe data”, *Proc. of ACM URBANE 2012*, *CoNext 2012 Workshop*.
25. Hohwald, H., Frias-Martinez, E. and Oliver, N. (2010), “User modeling for telecommunication applications: experiences and practical implications”, *Proc. UMAP*, pp. 327–338.
26. Kuzmenkova, S.V. (2014), “Subscribers` mobility as a potential factor of the cellular services consumption” *iCARE Junior, 2014,* HRU Higher School of Economics, Perm.
27. Rhee, I., Shin, M., Hong, S., Lee, K., Kim, S.J. and Chong, S. (2008), “On the levy-walk nature of human mobility”, *Proc. of INFOCOM.*, 2008.
28. Vieira, M.R., Frias-Martinez, V., Oliver, N. and Frias-Martinez, E. (2010). “Characterizing dense urban areas from mobile phonecall data: discovery and social dynamics”, *Proc. IEEE SocialCom, 2010,* pp. 241–248.
29. Wei, L., Zheng, Y. and Peng, W., (2012), “Constructing popular routes from uncertain trajectories”, *18th SIGKDD Conference on Knowledge Discovery and Data Mining*, 2012.
30. Zheng, K., Zheng, Y., Xie, X. and Zhou, X. (2012), “Reducing uncertainty of lowsampling-rate trajectories”, *International Conference on Data Engineering, 2012.*

Appendix 1

Tariff distribution

Tariff id	Frequency	Percent
40	569	12,32
37	552	11,95
20	463	10,02
42	439	9,50
43	428	9,27
53	313	6,78
51	302	6,54
34	283	6,13
41	281	6,08
1	225	4,87
52	214	4,63
47	134	2,90
8	78	1,69
16	77	1,67
21	53	1,15
49	31	0,67
45	30	0,65
39	23	0,50
17	22	0,48
7	21	0,45
11	18	0,39
14	10	0,22
22	10	0,22
50	10	0,22
10	7	0,15
48	5	0,11
4	4	0,09
18	4	0,09
3	3	0,06
5	3	0,06
23	2	0,04
2	1	0,02
6	1	0,02
9	1	0,02
19	1	0,02
29	1	0,02
Total	4550	100,00

Appendix 2

Spearman correlation coeffitients

	outcal~d	incall~d	distance	freque~y	distan~y	gyration	HH	fixed	mfee	free
outcalls_d	1.0000									
incalls_d	0.6816*	1.0000								
distance	0.5121*	0.4980*	1.0000							
frequency	0.6010*	0.5904*	0.8759*	1.0000						
distance_day	0.4923*	0.4629*	0.8986*	0.7842*	1.0000					
gyration	0.1503*	0.1328*	-0.0371*	0.1431*	0.0757*	1.0000				
HH	-0.2106*	-0.1799*	-0.5884*	-0.6092*	-0.5546*	-0.0739*	1.0000			
fixed	-0.1293*	-0.0968*	-0.1079*	-0.1030*	-0.0965*	0.0266*	0.0423*	1.0000		
mfee	-0.1326*	-0.1027*	-0.1272*	-0.1574*	-0.1265*	-0.0281*	0.1178*	0.0730*	1.0000	
free	0.1075*	0.0628*	0.0826*	0.0907*	0.0801*	-0.0092*	-0.0440*	-0.1997*	0.2384*	1.0000
c1	-0.2796*	-0.1846*	-0.1513*	-0.1906*	-0.1684*	-0.0454*	0.1052*	-0.0009*	0.2849*	-0.1156*
s1	0.2626*	0.1720*	0.2013*	0.2256*	0.1960*	0.0093*	-0.1033*	-0.0595*	-0.0244*	0.0900*
num_tariffs	-0.0079*	-0.0104*	-0.0241*	-0.0314*	0.0138*	0.0070*	-0.1200*	0.0083*	-0.1131*	-0.0146*
num_months	0.1735*	0.1325*	0.1168*	0.1298*	0.0817*	0.0887*	-0.0879*	0.0054*	-0.0887*	0.0209*
pack_d	0.2543*	0.1978*	0.2377*	0.2515*	0.2105*	0.0007*	-0.1179*	-0.3762*	-0.3006*	-0.0870*
mms_q_d	0.1734*	0.1727*	0.1027*	0.1315*	0.1178*	0.0546*	-0.0253*	-0.0161*	-0.0348*	0.0115*
gprs_q_d	0.1932*	0.1835*	0.1765*	0.2266*	0.1954*	0.1267*	-0.1012*	0.0078*	-0.0520*	-0.0004*
monthly_fe~y	-0.1946*	-0.1382*	-0.1672*	-0.1975*	-0.1648*	-0.0186*	0.1354*	0.2373*	0.9486*	0.1351*

	c1	s1	num_ta~s	num_mo~s	pack_d	mms_q_d	gprs_q_d	monthl~y
c1	1.0000							
s1	-0.1212*	1.0000						
num_tariffs	-0.1321*	-0.1186*	1.0000					
num_months	-0.0490*	0.0393*	0.0746*	1.0000				
pack_d	0.1937*	0.2480*	-0.1076*	0.0765*	1.0000			
mms_q_d	-0.0464*	0.0248*	0.0021*	0.0231*	0.0558*	1.0000		
gprs_q_d	-0.0483*	0.0656*	0.0267*	0.0342*	0.0827*	0.3512*	1.0000	
monthly_fe~y	0.4479*	-0.0582*	-0.1177*	-0.0844*	-0.2827*	-0.0393*	-0.0447*	1.0000

*-p<0,01 level of significance

Appendix 3

Comparison of estimations on 24 and 12 tariffs samples

Variable/Tariff id	37		20		42		43	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
Distance	-0,0006**	-0,0005913***	0,0002*	0,0002**	-0,0007***	-0,0007***	-0,0001	-,0001
Fixed fee	-0,115***	-0,1147612***	0,331***	0,331***	-0,021***	-0,021***	0,146***	0,146***
Monthly fee	-0,671***	-0,6710986***	0,390***	0,39***	-0,650***	-0,649***	0,0407***	,0407***
pc1*	69,99***	69,99353***	-176,1***	-176,12***	55,49***	55,487***	-153,5***	-153,4937***
ps1	-82,08***	-82,0826***	-198,2***	-198,167***	-13,57***	-13,567***	-177,2***	-177,244***
Number of tariffs	0,924***	0,9238816***	-0,201***	-0,2***	0,159	0,158	0,401***	0,4006***
Number of months	-0,082*	-0,0821263***	0,055***	0,054***	-0,124***	-0,124***	-0,0280	-0,0280435
Package	-3,160***	-3,160342***	-1,719***	-1,718***	-3,228***	-3,227***	-6,418***	-6,418***
MMS	0,253	0,252745	-0,0178	-0,0177	-0,160	-0,16	-0,0313	-0,031
GPRS	0,002	0,0025554	-0,0694	-0,069	0,424***	0,423***	0,406***	,406***
Constant	82,33***	82,33212***	387,1***	387,053***	-15,84***	-15,839***	376,3***	376,327***
Observations	46861	40821	46861	40821	46861	40821	46861	40821
Pceudo R-sq	0,524	0,98	0,524	0,98	0,524	0,98	0,524	0,98
AIC	127342,5	242	127342,5	242	127342,5	242	127342,5	242
BIC	130861,9	1284,7	130861,9	1284,7	130861,9	1284,7	130861,9	1284,7

Appendix 4

Discriptive statistics for different group of mobility levels

Tariff id	Obs	Mean	Std, Dev,	Min	Max
1*	1989	363,5803	400,0332	0	2700,63
41	2675	343,8672	383,4431	0	4728,118
51	2816	327,1661	329,7464	0	2781,872
20	5083	420,7307	435,6104	0	4071,744
34	2795	218,5129	283,9625	0	2457,388
37	5200	177,822	230,0238	0	3487,169
42	4097	236,3339	292,2921	0	3518,192
43	4249	403,4569	454,9783	0	6073,653
47	1363	365,0805	407,0591	0	2452,269

*Variables denote the level of mobility divided by9 tariffs