FACULTY OF ENGINEERING
AND ARCHITECTURE

Multiparameter analysis of roaming volumes in abstracted regions in context of the roam like at home legislation

Zicong Mai
Student number: 01300083

Supervisors: Prof. dr. ir. Sofie Verbrugge, Prof. dr. ir. Didier Colle
Counsellors: Jonathan Spruytte, Dr. ir. Marlies Van der Wee, Timo Latruwe

Master's dissertation submitted in order to obtain the academic degree of
Master of Science in Industrial Engineering and Operations Research

Academic year 2018-2019

GHENT
UNIVERSITY

FACULTY OF ENGINEERING
AND ARCHITECTURE

Multiparameter analysis of roaming volumes in abstracted regions in context of the roam like at home legislation

Zicong Mai
Student number: 01300083

Supervisors: Prof. dr. ir. Sofie Verbrugge, Prof. dr. ir. Didier Colle
Counsellors: Jonathan Spruytte, Dr. ir. Marlies Van der Wee, Timo Latruwe

Master's dissertation submitted in order to obtain the academic degree of
Master of Science in Industrial Engineering and Operations Research

Academic year 2018-2019

GHENT
UNIVERSITY

Acknowledgements

I would like to thank the people who have made this master's dissertation possible. First, I want to thank my supervisors Prof. dr. ir. Sofie Verbrugge and Prof. dr. ir. Didier Colle for offering me the opportunity to perform research on this interesting topic. Secondly, I want to express my sincere gratitude to my counsellors Jonathan Spruytte, Timo Latruwe and dr. ir. Marlies Van der Wee for providing me with necessary help and feedback for my work during the course of this research. They were always available and provided me with solutions and adjustments when problems occurred. Without them this dissertation would not have been possible. Finally, I want to thank everyone who supported me throughout my education. In particular my parents, for providing me the opportunity to study and their support throughout all these years.

Zicong Mai, August 2019

Copyright

Abstract

The European Commission introduced its roaming legislation in response to the overly expensive roaming prices at the time back in 2007. The term roaming refers to the use of mobile connectivity outside the end user's home network. The aim of the legislation was to encourage competition between mobile operators and end users. The latest step was the introduction of Roam Like at Home (RLAH) which removed the surcharges mobile operators could charge their customers for roaming in the European Economic Area (EEA). This led to large increases in consumed roaming volumes. The largest increase was observed in roaming data services. This made mobile operators question what the impact of RLAH will be on their network and costs. This master's dissertation aims to analyse the impact of RLAH in different countries/regions in the EEA. This was done by developing a conceptual model containing potential factors influencing the roaming use. The factors with available data were tested for their importance using a multiple regression approach. The results suggested that the regression model was unable to explain sufficient variability of the roaming use. This was potentially a result of the data limitations. As a consequence, no substantiated conclusions could be drawn related to the impact of RLAH in the different countries. The most promising parameters according to the results were: mobile penetration, retail price, and wealth per adult. Nevertheless, the factors included in the conceptual model are supported through literature. Therefore, the developed conceptual model contributes to the current roaming topic.

Multiparameter analysis of roaming volumes in abstracted regions in context of the roam like at home legislation

Zicong Mai

Supervisors: Prof. dr. ir. Sofie Verbrugge and Prof. dr. ir. Didier Colle

Counsellors: Jonathan Spruytte, Timo Latruwe, and dr. ir. Marlies Van der Wee

***Abstract* — The introduction of Roam Like at Home (RLAH) in 2017 has led to a large increase in mobile roaming use, i.e. the use of mobile connectivity in a foreign country. The largest increase was observed in roaming data use. This made mobile operators question what the impact of this evolution will be on their network and costs. This dissertation aims to analyse the impact of RLAH in different countries/regions in the European Economic Area. A conceptual model was developed containing the potential factors influencing the outbound roaming volumes. The factors with available data were tested for their importance using a multiple regression approach. The results suggested that the regression model was unable to explain sufficient variability of the roaming use. This was potentially a result of to the data limitations. The most promising parameters according to the results were: mobile penetration, retail price, and wealth per adult. Nevertheless, the factors included in the conceptual model are supported through literature. Therefore, the developed conceptual model contributes to the current roaming topic.**

***Keywords* — Roam like at home; Roaming; Conceptual model; Europe; Multiparameter analysis**

I. CONTEXT AND MOTIVATION

The European commission (EC) first introduced its roaming legislation back in 2007. The term roaming refers to the use of mobile connectivity outside the end user's home network. The realisation that there was insufficient competition in the European roaming market served as one of the main reasons for the creation of this legislation. The lack of competition led to the wholesale (prices operators can charge each other for the usage of their network) and retail (prices operators can charge end users) roaming prices being entirely decided by mobile operators. As a result, end users experienced overly expensive costs with the use of roaming. The roaming legislation aims to encourage competition in the roaming market and protect both end users and mobile operators by limiting the roaming prices by introducing price caps.

Due to lower roaming costs for mobile users together with the ever-increasing mobile evolution, roaming usage has increased heavily in the last years. The most recent step in the evolution of the legislation, i.e. the introduction of Roam Like at Home (RLAH) back in June 2017. With this new rule, operators in the European Economic Area (EEA) were no longer allowed to charge end users a higher price than the latter's domestic price when travelling within the EEA. Consequently, this has led to a vast increase in roaming usage. In the longer term, the assessment of the impact of RLAH on mobile operators requires an understanding of the factors influencing the roaming volume increases.

The rapid increase of roaming usage as a result of RLAH and its potential impacts on mobile operators serve as the main motivations for the research of this master's dissertation. The goal is to identify the influencing parameters and analyse the impact of RLAH on the increase in roaming volumes in different EEA countries/regions.

The goal of this research was to perform a multiparameter analysis on the mobile roaming volumes in abstracted regions in context of the RLAH legislation. This was done by developing a conceptual model containing the potential factors influencing the roaming volumes based on literature insights. A total of twelve potential factors were defined as input parameters in this model with the roaming use as the output. Next, each potential factor was tested for its importance using a regression approach. From these results, the potential impacts of RLAH on the different countries/region were estimated.

The structure of this paper is as follows. First, the basics of roaming necessary to understand the research topic are explained in section II. Then, the factors found is the literature are discussed in section III. Next, the conceptual model and the chosen parameters are summarised in section IV. In section V the results of the regression analysis are discussed. Finally, the conclusions regarding the important factors and the impact of RLAH on different countries are summarised in section VI followed by some future research topics in section VII.

II. BASICS OF ROAMING

In this section the basic concepts and terminologies used in roaming are explained.

A. *Costs associated with roaming*

There are several costs associated with roaming, these are summarised in Figure 1. To ensure that an end user can still access mobile services in a foreign country, its home operator, the DSP, needs to rely on the network of a foreign operator (FSP) to provide connectivity.

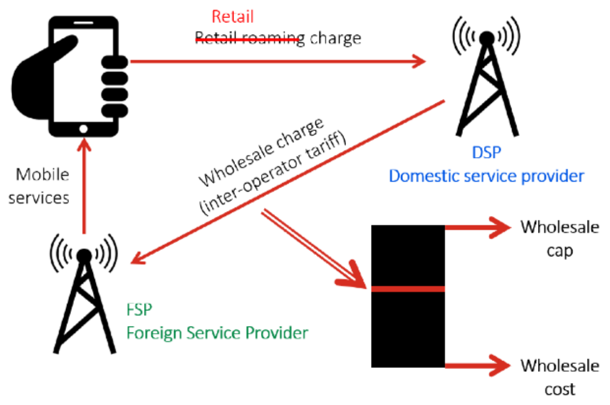

Figure 1: Roaming vs. RLAH [1]

In general, the service providers reached an agreement beforehand on what price should be paid when the end users of one operator use the network of the other. This is called the wholesale charge or inter-operator tariff (IOT). Before RLAH, the DSP could recover this cost by charging its customers with a retail roaming charge on top of the domestic retail price. This revenue is lost with RLAH since this new regulation removed this retail roaming charge completely for EEA mobile users roaming in the EEA.

Besides removing the retail roaming charges for end users, the EC also further revised the inter-operator tariffs for the wholesale roaming market. In this market, there are three different costs that are important to understand. First, there is the wholesale cost, i.e. the actual cost for the FSP to allow roamers' traffic on its network. Secondly, there is the wholesale cap, i.e. the maximum fee the FSP can charge the DSP for the use of its network by the DSP's customers, this is regulated by the EC. Finally, the wholesale charge or the inter-operator tariff is the actual price that the DSP pays to the FSP. Ideally, this charge lies between the wholesale cost and the wholesale cap [2]. Currently, this has not been achieved yet.

B. Operators and traffic types

It is also important to understand that there are different types of operators and traffic. In general, there are two types of operators. First, there is the Mobile Network Operator (MNO) which has its own network and radio spectrum, that can provide the full range of mobile services. Secondly, there is the Mobile Virtual Network Operator (MVNO). This operator differs from the MNO in the fact that it does not have its own radio spectrum [1]. They need to rely on MNOs to have access to it. Within the MVNO category, there are further differentiation possible based on the degree of dependence on the MNO. Some only need access to the radio spectrum while others may need the MNOs infrastructure as well to provide mobile services to their customers. Another way of dividing operators is based on their geographical coverage area. In this way, there are also two categories of operators: (1) the single country operator which only has network infrastructure in its home country, and (2) the cross-country operator which has network infrastructure in foreign regions as well.

In mobile roaming there are two types of traffic which are important from a mobile operator's point of view. The first one is the inbound traffic; this is, from a domestic operator's point of view, the mobile traffic that originates from foreign customers which needs to be handled on its own network. The second one is the outbound traffic; this is, again from a domestic operator's point of view, the mobile traffic that originates from its own customers which needs to be handled on a foreign operator's network.

From the concepts discussed above, we can already expect that analysing the impact of RLAH is quite complex since each type of operator will experience a different impact. For example. operators with more inbound compared to outbound roaming traffic will have to access whether their network can handle the traffic increase. On the other hand, operators with more outbound traffic will see an increase in total wholesale costs they need to pay. The different operator types will also experience a different impact which further complicates the analysis. The terminology explained in this section will be used throughout this paper; therefore, it is important for the reader to understand the differences.

III. Factors retrieved from literature

Since there are few studies focusing on forecasting roaming volumes specifically, existing studies related to (domestic) mobile internet use were analysed in order to identify potential factors of influence.

The factors summarised in this section are largely based on [3]. The authors of that paper reviewed 175 scholarly empirical publications on mobile internet (MI) usage intensity levels at individual subscriber level. They compared the results of these publications and discussed which factors were the most promising in explaining a user's mobile internet use intention. These are grouped together in the following three categories: (1) personal characteristics, (2) behaviour intention and attribute perceptions, and (3) factual use conditions

A. Personal characteristics

According to the analysis of [3], the most correlated factors are: <u>*Educational level*</u> and self-assessed general openness for innovative technical durables (<u>*Innovativeness*</u>). The factors in this category displayed a small to medium effects on the mobile internet usage intensity according to [4].

B. Behaviour intention and attribute perceptions

The factors in this category are related to one's intention to use mobile internet as well as a user's perceptions on MI. The largest correlation in this category was observed for <u>*Enjoyment*</u>, i.e. the hedonic or intrinsic value experienced when accessing the internet over a cellular data network. Furthermore, the behavioural intention to use MI and <u>*Ease of use*</u> perceptions also showed positive associations with MI usage. These factors tended to display larger mean effects on the mobile internet usage than the personal characteristics factors according to [3].

C. Factual use conditions

This category encompasses objective technical performance parameters of the network (e.g. transmission speed and latency) of a person's mobile service provider and of her access device (e.g. screen size and resolution, memory capacity). Here, it was found that a user's mobile internet <u>*Tariff plan*</u> and the <u>*Network quality*</u> had an important influence on the mobile internet usage [3].

IV. Conceptual model

Based on the factors retrieved from the literature review, a conceptual model is developed. The output variable of the model is a user's outbound roaming data use per time period. The choice of focusing on the outbound volumes is due to the limited available data. Since there was no data for the inbound volumes, these could not be analysed using the model. The input parameters of this model are inspired by literature insights, but these were extended with roaming specific parameters based on intuitive reasoning and the available datasets. The selected input parameters are the following:

- Education level
- Innovativeness
- <u>Income</u>
- <u>Travel frequency</u>
- Enjoyment
- Ease of use
- Usefulness
- <u>RLAH awareness</u>

- Mobile penetration
- Retail price
- Domestic data use
- Network quality

It can be seen from this list that there are several parameters which were not (directly) found in literature. The reason for this is because, as explained earlier, there few studies focused on roaming specifically; therefore, factors from literature were focused on mobile internet in general. It is easy to understand that additional parameters are needed, these are chosen intuitively based on the available datasets. The parameters in question are underlined in the list above.

The *Travel frequency* and *Roam Like at Home awareness* are added specifically to analyse the impact of RLAH since it is reasonable to assume that a higher travel frequency or awareness would have a positive influence on the roaming data use.

Mobile penetration rate is defined as the number of SIM cards in a country. The choice for using this parameter is based on the insights from the literature study. It was observed that the device type using the SIM card had an influence on the MI use.

The *Retail price* for using mobile data is added because this parameter is derived from the literature where it was found that the tariff plan of a mobile user and its monetary value were of influence on the MI use. The parameter incorporates both of those factors. Also, its potential interaction effect with the income parameter is expected to also influence the roaming volumes of the end user.

The *Domestic data use* serves as an indication for the roaming use, because it is expected that an end user's habit will carry over when travelling.

V. Results

A regression approach is chosen to fit the roaming volumes with the data of the input parameters. First, simple linear regressions are performed with each input parameter as the single predictor. The most promising predictors are selected by inspecting the regression coefficients, p-values and beta-values. Next, a multiple regression model is constructed with the promising predictors. The predictions are compared to the actual roaming volumes in order to draw conclusions about the impact of RLAH in different countries.

Data was collected from the available sources. For this research, only public databases were available. These are: BEREC roaming benchmark data report [5], Eurobarometer survey [6], EC's Digital scoreboard database [7] and Global wealth databook [8].

For some parameters in the conceptual model, there was no data available. Consequently, the effects of these parameters could not be estimated. Some intuitively chosen parameters were added as predictors in the regression in order to compensate for this problem.

The roaming use for each country is given as a monthly average per user. This meant that there was no differentiation possible between different user groups. Since the roaming use of a country is only represented by one value, the statistical power of the regression model is limited due to the limited number of samples. Also, only the outbound roaming data volumes were available from the databases; therefore, the decision was made to focus on the prediction of outbound volumes in this regression model.

It was also decided to only focus on the 28 member states of the EU instead of all countries in the EEA, because data for some countries were missing.

The parameters that were used as predictors for estimating the roaming use in the regression model are summarised in Table 1. It can be seen that the Turn off mobile data when travelling was also added as predictor for the regression model.

The original values of these parameters were converted into values relative to the EU-28 average before using them in the regression model, because this allowed an easier interpretation of the results.

Table 1: Parameters included in the initial regression model

Parameter
Roaming data use (outbound) [5]
Domestic data use [5]
Travelling in the EU-28 [6]
Wealth per adult [8]
Usefulness [6]
Turn off mobile data when travelling [6]
Mobile penetration rate [7]
Retail price [9]
RLAH awareness [6]

The results of the regression model are discussed in the following subsections.

A. Pairwise comparisons

Each of the parameters are tested individually as predictor (= independent variable) for the (outbound) roaming data use (= dependent variable). This is equivalent as eight simple linear regressions. The results are summarised in Table 2.

Table 2: Results of the pairwise comparisons

Predictor	Regression coefficient	p-value	beta-value
Domestic data use	0.2180	0.1430	0.3078
Travelling in the EU-28	0.2322	0.1833	0.2811
Wealth per adult	0.1685	0.1019	0.3420
Usefulness	0.0336	0.9641	0.0097
Turn off mobile data	0.1156	0.6054	0.1110
Mobile penetration rate	0.8156	0.0062	0.5424
Retail price	-0.4038	0.0111	-0.5091
RLAH awareness	0.0730	0.9128	0.0236

The signs of most coefficients in these pairwise comparisons make sense, i.e. the correlation relationships between the predictor and the output variable seems reasonable.

Based on the p-values, it can be seen that Mobile penetration (0.0062), Retail price (0.0111), and Wealth per adult (0.1019) seem to be the statistically significant predictors. However, the limited number of samples used in this regression model was insufficient to achieve reliable results. As a consequence, even if a predictor is significant in reality, this model could potentially fail to detect it.

From comparing the beta-values, it can be seen that the most significant predictors: Mobile penetration and Retail price, also show the largest impact on the output variable with beta-values of 0.5424 and -0.5091 respectively, followed by Wealth per adult (0.3420).

In theory, only these three predictors (Mobile penetration, Retail price, and Wealth per adult) should be included in the multiple regression model. However, due to the potential

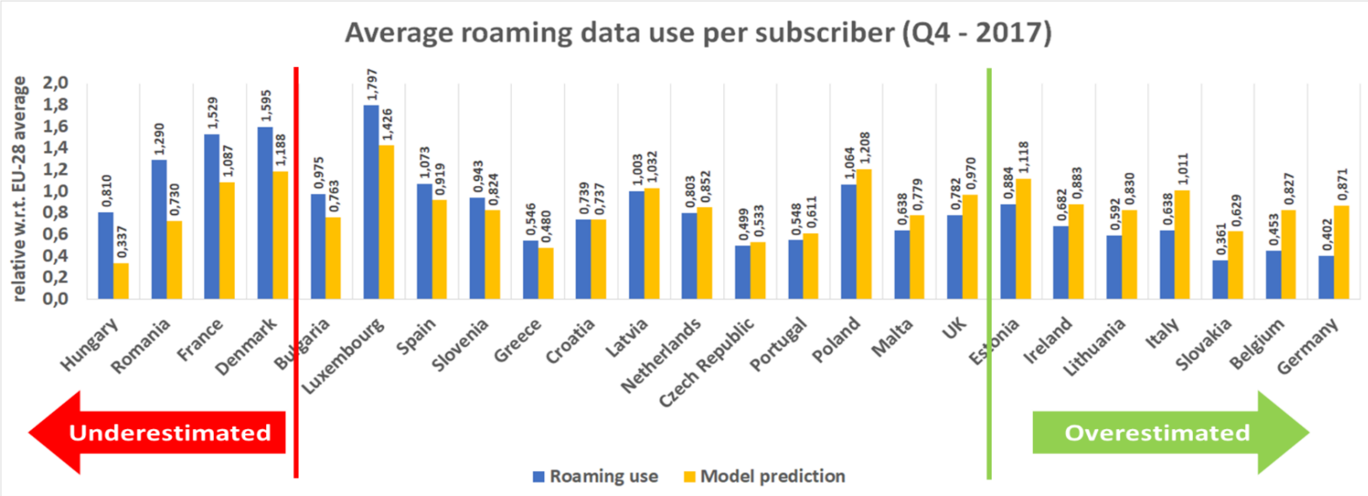

Figure 2: Comparison of model predictions and the actual roaming use

problems with data limitations, three other predictors which were also assumed to be of importance were also selected, these are: Domestic data use, Travelling in EU-28, and RLAH awareness. The reason for this choice was because some parameters, which are important in reality, could be missed out by the model, i.e. not achieving statistical significance based on the p-values, due to the data limitations. Therefore, one must not solely focus on the p-values, but use some intuitive reasoning as well for selecting parameters. Hence, some parameters, were kept in the multiple regression model despite them not showing statistical significance in the pairwise comparisons earlier on.

A. *Multiple linear regression*

From the results of the simple linear regressions, two of the initial parameters in Table 1 were not selected for the multiple linear regression model, i.e. usefulness and turn off mobile data.

The results of the initial model showed that the signs of most predictors' regression coefficients remain the same compared to the pairwise comparisons, except for Domestic data use and RLAH awareness. These two showed a negative correlation between with the Roaming use which seems counter-intuitive. The reason for this could lie in the fact that some other potential predictors are missing, e.g. interaction effects between current predictors. However, it was chosen to not further investigating the construction of these interaction terms because there was insufficient data to verify if these were good ones to include or not within the available time frame of this research.

In an attempt to obtain predictions in which the contribution of each parameter could be explained more rationally, the two predictors (RLAH awareness and domestic data use) were removed from the model. The updated model now only contains four predictors and the new results are summarised in Table 3.

Table 3: Multiple linear regression results

Predictor	Regression coefficient	p-value	beta-value
Intercept	0.4883	0.289	-
Mobile penetration	0.4962	0.136	0.3300
Retail price	-0.2397	0.171	-0.3023
Wealth per adult	0.1275	0.243	0.2589
Travelling in the EU-28	0.0098	0.957	0.0118

The predictors' regression coefficients and signs make sense. Mobile penetration has the largest coefficient value and will have a larger impact on the predictions since it absolute value (0.4962) is at least two times larger than the other predictors. It is remarkable that the influence of the travelling predictor is much smaller compared to the simple linear regression results, there it was found that it had a larger regression coefficient and beta-value. However, its effect might be lowered when used together with other predictors like in this model.

None of these predictors seem to be statistically significant according to the p-values, but mobile penetration is still the closest to being statistically significant.

The signs of the beta-values correspond to those of the regression coefficients which is expected. The absolute beta-value of mobile penetration is decreased while that of the retail price is increased compared to the previous model making their impact on the roaming use more similar.

The main takeaways from these results are: (1) potential interaction effects could have more impact on the roaming volumes than individual predictors and (2) the most promising parameters, based on these results, are: Mobile penetration, Retail price, and Wealth per adult.

The predictions of this multiple linear regression model are compared the actual values of roaming use. One must also note that the predictions here are produced by only four predictors (mobile penetration, retail price, wealth per adult, and travelling in EU-28) which means that the potential impact of other parameters are neglected. The comparisons are shown in Figure 2

The countries on the left side of this figure are the ones where the roaming use is underestimated by the multiple regression model, while the countries on the right side are overestimated. It was decided to label a prediction acceptable when its deviation from the actual value stayed below 25%. Using this criterium, eleven out of the twenty-four countries considered in the regression analysis were either under- or overestimated. The underestimated countries do not share specific characteristics, the same holds for the overestimated ones. For example, in the underestimated countries, there are the less wealthy countries (Hungary, Romania), but also very wealthy countries (France and Germany). Also, their retail prices and mobile penetration, which have the most impact on these predictions, also vary a lot. The same observations can be made in the overestimated countries. Consequently, it was not possible to group countries together with similar characteristics or define abstracted regions based on the results of this model.

It can thus be concluded that this regression model is unable to explain sufficient variation in the roaming use. This reinforces the belief that interaction effects of parameters could also influence the roaming use, e.g. when an individual is wealthier, then that person could be less affected by the retail price and still chooses to roam more regardless of the higher price. Hence, more complex forecasting methods may be necessary to model these interaction effects.

VI. Conclusions

The data limitations limited the statistical power of the regression model. As a result, parameters which are important in reality could potentially be overlooked by the model. Therefore, the conclusions in this subsection must still be examined more thoroughly once more data becomes available.

The developed conceptual model in this dissertation only focused on analysing outbound roaming data volumes. The literature input was also limited since few existing studies focused on analysing potential factors of influence on the roaming use. This is partially understandable since roaming data use in the EEA only started to increase exponentially from 2017 onwards due to RLAH. Hence, the available datasets also served as inputs for the model parameters.

An initial data analysis was performed to detect potential outliers, this resulted in the removal of four countries (Cyprus, Sweden, Austria and Finland) from the regression analysis.

The simple linear regressions showed that only the following parameters were statistically significant based on their p-values: (1) Mobile penetration, (2) Retail price, and (3) Wealth per adult. However, three other parameters were selected to be used in the multiple regression model as well, these were: Domestic data use, Travelling in EU-28, and RLAH awareness.

The multiple linear regression model with these six parameters showed some unexpected results, i.e. the sudden change to a negative correlation between the roaming use and the predictors domestic data use and RLAH awareness. This suggested the presence of potential interaction effects between these predictors. Due to data limitations and the available time frame, it was difficult to perform a quantitative analysis on these interaction effects. Therefore, it was chosen to not further investigate the construction of these interaction terms because there was insufficient data to verify if these were good ones to include or not. These two predictors were excluded in the updated model and the new results showed that none of the predictors are now statistically significant. However, the most promising parameters are: Mobile penetration, Retail price, and Wealth per adult. These correspond to the results found in the simple linear regressions earlier on.

When the linear relationship assumption between the predictors and the roaming use of the regression model was examined, it was observed that a non-linear relationship for some of the predictors was more likely the case. Hence, a non-linear fitting approach might be a better option. However, this was not further investigated due to three reasons. First, a non-linear regression approach requires the input of the estimated relationships between the fitted variable and its predictors, e.g. linear, quadratic, cubic, etc. This additional estimation is already a complex process since the actual important parameters have not been known yet at this point. Together with the fact that the limited number of samples would not give reliable results, it seemed that this would further complicate an already difficult problem. Secondly, the focus of this research was to identify potential influencing factors using a more conceptual approach instead of purely data focused. Thirdly, other forecasting methods, e.g. time-series or machine learning, focus on one dataset only, i.e. the roaming volumes, and try to fit this dataset the best way possible instead of identifying the potential relationships between the roaming volumes and the chosen predictors (which is what is actually desired in this research). For these reasons, it seemed more useful to use the simpler multiple linear regression method because the predictors could be inserted and tested without further inputs that a non-linear regression model required. In this way, an initial view on the potential important parameters could be obtained.

The predictions of the multiple regression were then compared to the actual roaming volumes. A difference of less than twenty-five percent was assumed to be acceptable. There was no clear differentiation observed in the countries that were either over- or underestimated, e.g. it could not be said that less wealthy countries tend to be underestimated by this model. The same holds for the countries with acceptable predictions. As a result, grouping countries together or defining abstracted regions containing countries with similar characteristics were not possible. Eleven out of the twenty-four countries considered in this model were either over- or underestimated. Consequently, this regression model proved to be insufficient in explaining the variability of the roaming volumes, which was expected due to the potential non-linear relationship between some of the predictors and the roaming use as observed from verifying the model assumptions.

With the aim of analysing the impact of RLAH for different operators and regions, the question raises whether the forecast should focus on the average roaming use instead of peak use in different time periods or seasons. The average value might be misleading since it conceals the very heavy users from the normal users which is currently the case with only one data point per country. This is a potential problem because according to the literature, a small part of the users (i.e. the heavy ones) are responsible for the majority of the data use. Consequently, special attention should be paid for these users from an operator's perspective. It could be more meaningful for operators to focus on the evolution in roaming use for different user groups (i.e. light, normal, heavy) in order to obtain a more accurate view on which part of users will most likely be responsible for the largest increase in roaming use and by extension the change in costs. Also, different types of operators will experience a different impact, e.g. operators from net receiving countries such as Spain are more interested in the evolution of the inbound volumes since this is the majority of their roaming traffic. For these reasons, it is highly unlikely that one model will be sufficient to model the complex phenomenon of roaming use. Therefore, further differentiation of user groups and traffic types are necessary to forecast the roaming volumes more accurately. The models for each user group might contain different input parameters because each group has different characteristics. The same goes for the inbound versus the outbound models.

It can thus be concluded that the mobile roaming use, especially data services, is a complex phenomenon which cannot be analysed thoroughly without sufficient and reliable datasets. Hence, more complex forecasting methods, e.g. non-linear approaches, and further differentiation between user groups and traffic types might be necessary to better model the roaming use. Nonetheless, this research contributed to the roaming topic by developing a conceptual model with the parameters supported by literature. The results of the limited regression model suggested the potential importance of three parameters. It also reinforced the belief that using one model will most likely not be sufficient. The insights gained in this research and the future work suggestions in the next section serve as a good starting point from which future researchers can perform a better analysis on roaming and obtain better estimates without having to start from scratch. As a result, a better view on the evolution of roaming use in different regions and countries can be obtained from which the real impact of RLAH in terms of cost changes for mobile operators can be seen.

VII. Future work

In this section, some interesting future work topics building on the insights gained through this research are given.

First, more data samples must be collected from a representative population, i.e. users with different user profiles, of all ages, etc. This is necessary in order to obtain more reliable results. To achieve this, an advanced collaboration between mobile operators is needed to provide enough data samples for each parameter of the conceptual model.

Secondly, a number of studies in the literature observed that data use across mobile customers are highly skewed, i.e. a very small portion of "heavy" users causes a large fraction of the mobile data carried over the mobile operator's network. Therefore, it would be useful to split the users into three types/categories: 'light', 'medium', and 'heavy' users. In this way, a separate model for each type of users will result in better estimates for each group, because each parameter will have a different coefficient/impact for each type of users. This was, e.g. something that could not be investigated in this dissertation because the roaming volumes were only available as a monthly average per subscriber. The average values will even out the peak volumes consumed by the 'heavy' users.

Next, different models can be explored as well. The current model is focused on outbound volumes, the logical next step would be to apply the model to analyse the inbound volumes as well. When both volumes are available, a better view can be obtained on which countries are net receiving or net sending.

In addition, the total roaming volumes can be analysed as well, e.g. a model can be fit to the total volumes of a certain country in a timespan (monthly, quarterly, yearly, etc.). It would be interesting to compare forecast of the total volumes (forecasted a whole) with the sum of the separate forecasts of the in- and outbound volumes. In this way, it can be examined whether forecasting the in- and outbound volumes separately will give more accurate predictions compared to the forecast of the total roaming volumes as a whole.

Besides these general models, it could also be useful for operators to focus their attention on the peak volumes instead of the total yearly volumes or per subscriber use. The reason for this is because operators from net receiving countries will want to know if their network is able to handle the traffic spikes during a very busy period or a specific season where there are more tourists visiting the country.

The models in the previous paragraphs fitted the roaming volumes with data from different potential influencing factors, however, it can also be useful to fit the roaming volumes without these factors by using techniques that only looks at one dataset, i.e. the roaming use, and try to determine a function that fits this data in the best possible way. These predictions can be compared to the ones obtained from the models discussed above to see whether predictions using multiple factors as input parameters are better than predictions using only the roaming volumes as input.

Up till now, the focus was on estimating the roaming volumes' evolution. However, to analyse the real impact on operators this is only the first step. The roaming volume estimates are used as input in the cost model of operators. This cost model is what is important for mobile operators. In such a model, different inputs, other than the roaming volume predictions, are used to estimate the total cost of providing roaming services for their subscribers; these include infrastructure costs, wholesale costs, other business operating costs, etc. From the output of this cost model, the actual change in costs (as a result of RLAH) can then be analysed. This is the real impact operators will experience.

Lastly, since other parts in the world (Asia, South America, Africa, etc.) are also evolving into a roaming like at home principle (using the EEA as a benchmark) in their respective regions, it would be interesting to use the models with the coefficients based on EEA countries to predict roaming volumes in other parts of the world to see whether the same accuracy can be achieved. Additionally, the usefulness of each input parameter can be tested and verified for other regions because it is possible that some parameters, which were valid for the EEA, are not applicable anymore to another region.

References

[1] F. Lagaet, "Economic implications of reducing or abolishing roaming fees," 2016-2017.

[2] J. Spruytte, M. Van der Wee, M. Regt, S. Verbrugge and D. Colle, "International roaming in the EU: Current overview, challenges, opportunities and solutions," *Telecommunications Policy,* vol. 41, pp. 717-730, 2017.

[3] T. J. Gerpott and S. Thomas, "Empirical research on mobile Internet usage: A meta-analysis of the literature," *Telecommunications Policy,* vol. 38, pp. 291-310, 2014.

[4] I. Lee, J. Kim, B. Choi and S.-J. Hong, "Measurement development for cultural characteristics of mobile Internet users at the individual level," *Computers in Human Behavior,* vol. 26, pp. 1355-1368, 2010.

[5] BEREC, *International Roaming BEREC Benchmark Data Report April 2017 - September 2017,* 2018.

[6] E. Commission, *Flash Eurobarometer 454: The end of roaming charges within the EU,* 2017.

[7] E. Commission, *Digital Scoreboard - Data & Indicators,* 2018.

[8] C. Suisse, *Global Wealth Databook 2017,* 2017.

[9] E. Commission, *Mobile Broadband Prices in Europe 2017,* 2017.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ARIMA	Autoregressive integrated moving average
ARPU	Average revenue per user
BEREC	Body of European Regulators for Electronic Communications
CITEL	Inter-American Telecommunication Commission
DG	Directorate-General
DSP	Domestic Service Provider
EAC	East African Community
EC	European Commission
ECOWAS	Economic Community of West African States
EEA	European Economic Area
EU	European Union
EU-28	28 member states of the European Union
FSP	Foreign Service Provider
GSMA	Global System for Mobile Communications
ICT	Information and Communications Technology
IDT	Innovation Diffusion Theory
IMR	International Mobile Roaming
INTUG	International Telecommunications Users Group
IOT	Inter-Operator Tariff
ITU	International Telecommunication Union
MI	Mobile Internet
MNO	Mobile Network Operator
MVNO	Mobile Virtual Network Operator
OECD	Organisation for Economic Cooperation and Development
PEOU	Perceived Ease Of Use
PU	Perceived Usefulness
QoS	Quality of Service
RLAH	Roam Like At Home
RoW	Rest of the World
SADC	Southern African Development Community

SIM	Subscriber Identification Module
SMS	Short Message Service
STIRA	Standard Term for Internation Roaming Agreements
TAM	Technology Acceptance Model
TPB	Theory of Planned Behaviour
UTAUT	Unified Theory of Acceptance and Use of Technology
VNI	Visual Network Index

Chapter 1

Context and motivation

The European commission (EC) first introduced its roaming legislation for European countries in 2007 in response to the overly expensive roaming prices back then. Roaming refers to the use of mobile connectivity in a foreign country. The reason for the legislation was the realisation that there was insufficient competition in the European roaming market. Due to this, the wholesale (prices operators can charge each other for the use of their network) and retail (prices operators can charge end users) roaming prices were entirely decided by mobile operators. This led to overly expensive costs for mobile users when they were using mobile services in a foreign country which resulted in the so called 'bill shock'. This term refers to mobile users receiving very expensive bills for often only a small amount of roaming use. The European roaming legislation aims to encourage competition in the roaming market and protect both end users and mobile operators by limiting the roaming prices (both wholesale and retail) by introducing price caps. In the beginning, only voice service prices were capped, but caps for text services (SMS) soon followed in 2009. In the same year, wholesale caps for data services were also introduced. The retail caps for data were only introduced in 2012. This legislation has passed many revisions since its inception in 2007. Both wholesale caps and retail caps have been altered multiple times in the last decade. These caps are continuously being reviewed in order to adapt to the market situation, e.g. the current caps for voice and text services are only valid until the end of 2019, the new caps for the next few years are yet to be announced by the EC.

Due to lower cost of roaming in combination with the ever-increasing mobile evolution, roaming usage has increased heavily in the last years. The most recent step in the evolution of the legislation was the introduction of the Roam Like at Home initiative (RLAH) in June 2017. This says that mobile operators were no longer allowed to charge end users a higher price than the latter's domestic price for using mobile services when travelling within the EEA. Consequently, this has led to a vast increase in roaming usage. The increase for data services was the largest, this was already observed in the first quarter after the introduction of RLAH.

In the longer term, the assessment of the impact of RLAH on mobile operators requires an understanding of the factors influencing the roaming volume increases. Once these factors are identified, mobile operators can use these to estimate how the roaming volumes will evolve. Then, they can assess whether their current tariff plans need to be adapted or whether their network infrastructure needs to be upgraded to handle the traffic increases. The rapid increase of roaming usage due to RLAH and its potential impacts on mobile operators serve as the main motivations for the research of this master's dissertation. The goal is to analyse the impact of RLAH on the increase in roaming volumes in different EEA countries/regions. This will be done by developing a conceptual model containing the potential factors influencing the roaming use. These factors are then tested for their importance using a regression approach to analyse the impact of each factor. From these results, the impact of RLAH on a specific country is estimated.

The methodology used in this dissertation is the following. First, the necessary background related to roaming and the evolution of the its legislation will be obtained from a literature study. Secondly, existing studies related to mobile (roaming) use will be analysed to gather potential influencing factors. Thirdly, a conceptual model will be constructed based on the influencing factors found in the literature study. Fourthly, data for the model will be collected from the available sources. Next, a regression analysis will be performed with the collected data to obtain a fit for the roaming volumes per country. The obtained results will then be analysed in the context of the impact of RLAH on different countries. The focus of the model will be on identifying potential factors influencing the roaming data volumes instead of the other two services voice and text. The reason for this choice is because data services have the largest contribution to a mobile operator's network load. Data services use far more traffic compared to voice and text services.

The course of this master's dissertation is as follows. Chapter 2 explains the basic terminology related to mobile roaming followed by the historic evolution and the current trends of the European mobile roaming market. Next, a literature review is done in Chapter 3 on the existing studies related to mobile internet use from which the potential influencing factors are summarised. Chapter 4 then provides a detailed explanation of the conceptual model and how the input parameters are chosen based on the insights of the literature review. Chapter 5 summarises the collected data together with the results of a multiple linear regression analysis using the available data. Lastly, Chapter 6 summarises the premise, the obstacles, and the results of this research together with some guidelines for future research in this topic.

Chapter 2

Historic overview of roaming and current trends

In this chapter an overview of the basics of roaming, its evolution in different parts of the world, and the current trends will be given. This is necessary to understand the context and research question of this master's dissertation.

2.1 Basic roaming concepts

Roaming is a term in wireless telecommunication used to describe the ability of an end user to use cellular services (voice, text, data), when travelling outside the geographical coverage area of the home network, by means of using a visited network. The home operator of the end user, i.e. the domestic service provider (DSP) and the operator that manages the visited network called the foreign service provider (FSP).

In mobile roaming there are two types of traffic which are important from a mobile operator's perspective, these are shown in Fig. 2.1. The first one is the inbound traffic, i.e. from a domestic operator's perspective, the mobile traffic that originates from foreign end users which needs to be handled on its own network. The second one is the outbound traffic, i.e. again from a domestic operator's perspective, the mobile traffic that originates from its own end users which needs to be handled on a foreign operator's network [1].

2.1.1 Costs associated with roaming

There are several costs associated with roaming, these are summarised in Fig. 2.2. To ensure that an end user can still use mobile services outside the geographical coverage area of its home network, its home operator, the DSP, needs to rely on the network of a foreign operator

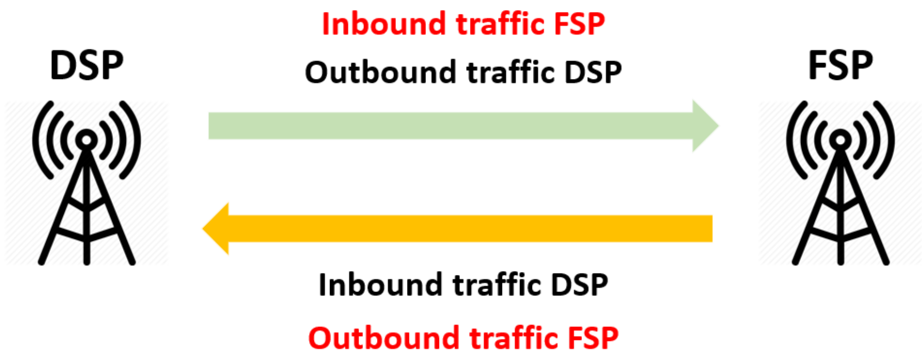

Fig. 2.1: Inbound and outbound traffic

(FSP) to provide connectivity[1]. In general, the service providers have reached an agreement beforehand on the price that should be paid when the end users of one operator use the network of the other. This is called the wholesale charge or inter-operator tariff (IOT). Before RLAH, the DSP could recover this cost by charging its end users a retail roaming charge on top of the domestic retail price. The roaming retail charge is the price an end user pays to the DSP for using of roaming services.

In the context of the Roam Like at Home (RLAH) initiative, this is specific for the European Economic Area (EEA), the DSP cannot charge its customers with a retail surcharge anymore since July 2017. This is part of the European Commission's roaming legislation. Besides removing the retail roaming charge for end users, the EC also further revised the inter-operator tariffs for the wholesale roaming market. In this market, there are three different costs that are important to understand. First, there is the wholesale cost, i.e. the actual cost for the FSP to handle roamers' traffic on its network. Secondly, there is the wholesale cap, i.e. the maximum fee the FSP can charge the DSP for the use of its network by the DSP's customers, this has been set by the EC. Finally, the wholesale charge or the inter-operator tariff is the actual price that the DSP pays to the FSP which is negotiated beforehand. Ideally, this fee lies between the wholesale cost and the wholesale cap. Currently, this has not been achieved yet.

The balanced/unbalanced pricing model

Negotiations on bilateral wholesale agreements between operators are based upon a set of different pricing models (fixed rate, balanced/unbalanced pricing, volume commitment, etc.). One of the more important pricing models is the balanced/unbalanced pricing model. In this kind of agreement, two operators agree to send traffic over each other's network. If both

[1]Except when the DSP is a cross-country operator, i.e. a service provider who owns network infrastructure in foreign countries

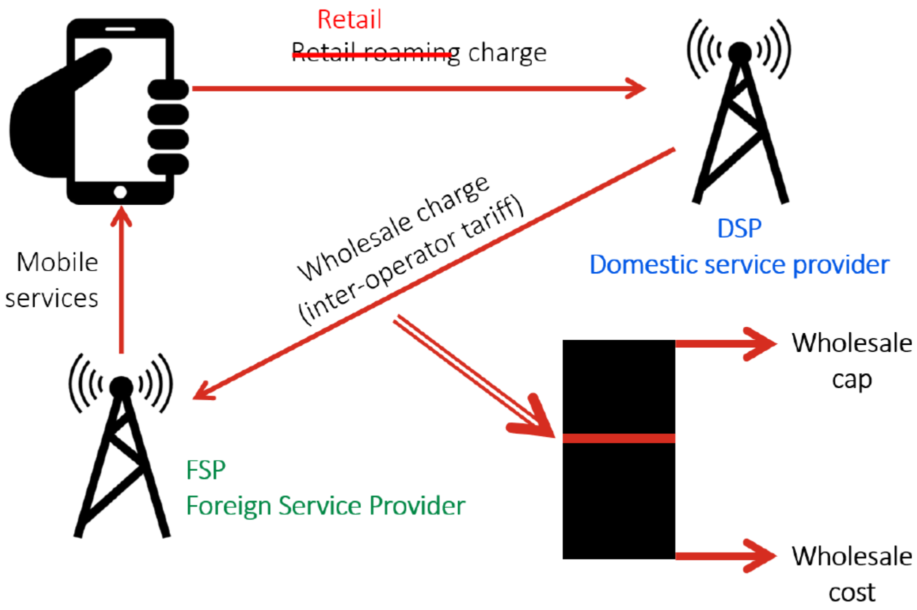

Fig. 2.2: Costs associated with roaming before and after Roam Like at Home [1]

operators send an equal amount of traffic, the exchange is balanced and so are the costs and revenues of each operator, resulting in a financial zero-sum game. When the amounts are unbalanced, the net sender operator (the operator with more outgoing traffic) pays a pre-discussed wholesale rate to the net receiving operator [2]. In this model, operators can thus have wholesale costs which are balanced or unbalanced. It can be seen that operators will benefit if they can approximate the balanced situation as good as possible.

2.1.2 Different types of operators

In the current market there are several types of telecom operators who offer mobile (roaming) services to its customers. These operators can be categorised in different ways depending on the used criteria. Two categorisations, which are most frequently used in literature, are discussed below.

In general, there are two types of operators. The first type is the Mobile Network Operators (MNOs). These operators have its own network, which consists of the actual network infrastructure and radio spectrum, that can provide the full range of mobile services, i.e. voice, text (SMS) and data. The second type is the Mobile Virtual Network Operators (MVNOs). They differ from MNOs in the fact that they do not have its own radio spectrum [1]. They need to rely on MNOs to have access to it. Within the MVNO category, there are several types that can be defined. They vary in the degree of dependence on the MNO. Some only need access to the radio spectrum while others may need to rely on the MNOs network infrastructure as well to provide mobile services to their customers. The most important types of MVNOs are summarised in Fig. 2.3. It can be seen that the even the Full MVNO still needs to rely on an MNO for the use of its radio spectrum in order the operate.

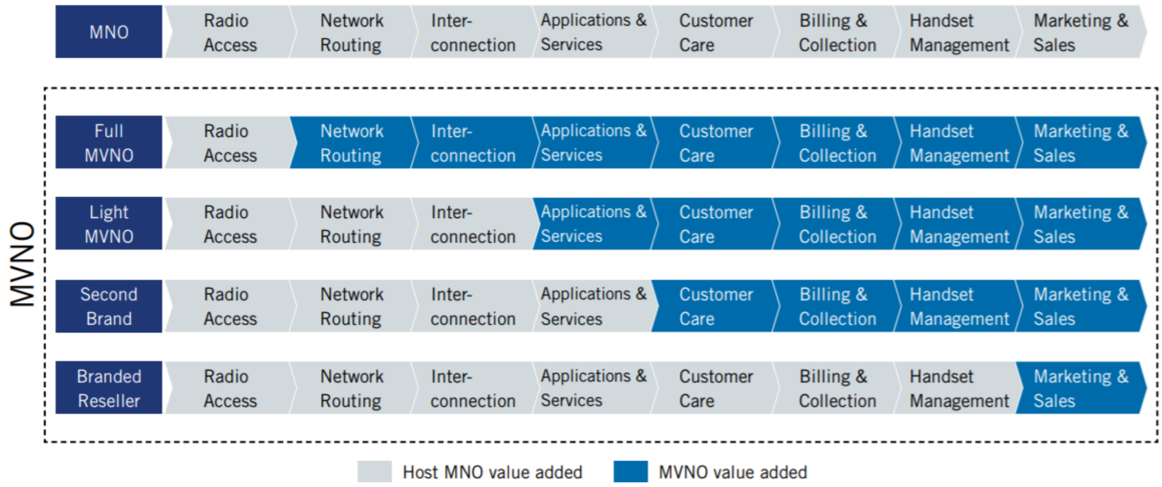

Fig. 2.3: Different types of operators [3]

Another way of categorising the operators into groups, is based on their area of coverage. Using this criterium, the operators can again be divided into two groups. The first one is the single country operators. Their area of coverage is limited within its own country, when roaming traffic of their customers needs to be handled, they have to rely on the services of an FSP. The second group consists of the cross-country operators. They typically have infrastructure outside its home country meaning that their geographical coverage is increased. In this way, they can let their customers roam on their own network by steering the traffic through its own network. The Deutsche Telekom in Europe is a good example of such a cross-country operator. They are active in thirteen countries. This means that they can avoid paying wholesale charges to FSPs in these countries. Consequently, they can offer more attractive tariff plans to its customers compared to its competitors [1].

2.2 Evolution and developments of the international mobile roaming market

In this section, the evolution in the international roaming markets will be described for different regions around the world. The international roaming market is not a single one, therefore, the progress of evolution in each region is different. For each region, some examples will be given to illustrate the current market situation. For a more in-depth analysis of the evolution and comparison of developments in different regions the following papers can be consulted: Sutherland [4], Bourassa, Paltridge, Weber, *et al.* [5] and ITU [6].

2.2.1 European Economic Area

The first European mobile roaming agreement between Mobile Network Operators (MNOs) was signed in 1992 between Vadofone UK and Telecom Finland [4]. As roaming agreements expanded, the Global System for Mobile Communications (GSMA), the trade body that represents the interests of mobile network operators worldwide, proposed a framework to simplify MNO roaming negotiations, namely the Standard Term for Internation Roaming Agreements (STIRA). However, this framework had the effect of suppressing competition and discounting, because it treated all MNOs and their price terms equally under its principle of non-discrimination. The retail price, that mobile users had to pay then, was calculated as follows: the wholesale price that the DSP had to pay the FSP plus up to 15% of that price as its retail margin. This meant that the retail and wholesale pricing were linked. However, the Directorate-General for Competition (DG competition) did not consider the wholesale or retail pricing to be cost based. Therefore, it modified its framework under the new name of Inter-Operator Tariff (IOT) scheme in 1997. This new scheme had some unintended consequences as well. Wholesale market prices were effectively free to float, so they increased a considerable amount. Consequently, the retail prices followed as well, because both were linked. A comparison between retail prices of the previous framework and the new one in Q4-2000 showed increases over 212% for peak and off-peak voice calls between EU member states for certain MNOs [7]. The International Telecommunications Users Group (INTUG) complained that the prices for roaming within the European Union (EU) were unjustifiably high and that competition was not bringing them down [8]. The operators offered a variety of explanations such as claiming that international mobile roaming (IMR) was a premium service justifying high prices, comparable to a bottle of wine in a restaurant. In response to these market conditions, the European Commission began their ten-year plan in 2007 in order to remove roaming surcharges within the EEA. The EEA consists of the 28 member states of the European Union plus Iceland, Liechtenstein and Norway. Switzerland is neither an EU nor EEA member. This roaming regulation is subdivided into four transition stages to gradually decrease the wholesale and retail prices [9].

As stated in the international roaming data benchmark report of the Body of European Regulators for Electronic Communications (BEREC) [10], the 2007 regulation was called Roaming I and it introduced the concept of Eurotariffs which placed caps on both the wholesale and retail prices for incoming and outgoing voice services. However, the roaming outside the EEA remained unregulated. The second roaming regulation (Roaming II) came into force back in 2009 and introduced the concept of glide path decreases and price regulation of data roaming services at wholesale level. The existing caps for voice services continued to decrease while price caps were also introduced for text services. To avoid bill shocks related to roaming data use, operators were required to notify their customers if their uses exceeded the amount of €50 excl. VAT. The mobile users can then decide whether they want to continue using roaming data or not. The third regulation (Roaming III), introduced back in 2012, further extended the existing caps for all mobile roaming services by finally adding retail caps for mobile data services as well. The caps introduced in the previous regulations served as preparation for the removal of retail roaming surcharges in the EEA. This new concept is called Roam Like at Home (RLAH) initiative. In 2014, the regulation was revised once again which resulted in Roaming IV. There it was decided that the existing wholesale caps would remain until 30 June 2022 while the retail surcharges would be removed on 30 June 2017. From then on, mobile operators can only charge its subscribers domestic rates for the use of roaming services. It must be mentioned that the wholesale caps defined back then will not remain until 2022. The EC is currently reviewing the caps for voice and text services for the next few years, these have not been announced yet. A visual representation of the glide path of Eurotariff caps on voice services is shown in Fig. 2.4 where the gradual decrease in caps can be observed.

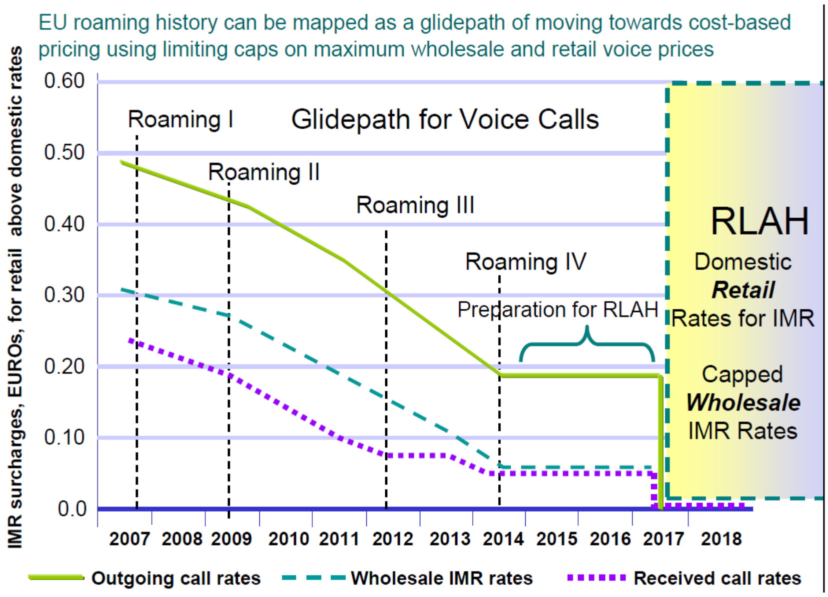

Fig. 2.4: Glide path of EU tariff caps for voice roaming 2007-2017 [9]

In order to prevent the abuse of RLAH, some measurements were taken to protect the mobile operators. RLAH is intended for people who occasionally travel outside the country where they live or have stable links, i.e. if they work or study there. It is not meant to be used for permanent roaming[2]. As long as you spend more time at home than abroad, or you use your mobile phone more at home than abroad, you can roam freely at domestic prices when travelling anywhere in the EEA. This is considered a "fair use of roaming services" and is incorporated in the fair use policy (FUP) [11]. Currently, there are no volume restrictions in

[2]Permanent roaming refers to the situation where an end user buys a SIM card from a foreign operator that offers lower pricing than any domestic operator and thus take advantage of the cheaper pricing and use "roaming" while also being at its home country [2]

RLAH for voice calls and text, but there are rules and limits for data use at domestic pricing which are determined by the type of contract the end user has. Mobile operators may apply fair, reasonable and proportionate control mechanisms to avoid abusive use of RLAH. When an abuse of roaming data use is detected, the operator can surcharge the abuser for it.

The EU's wide-ranging initiatives on mobile roaming prices are unique within the Organisation for Economic Cooperation and Development (OECD) and globally. Other countries have followed the EC's vision and started initiatives aiming at reducing prices and working towards an RLAH principle in their own regions [5].

2.2.2 America

South American countries have also taken initiatives to reduce roaming prices for their end users when travelling in the region. For example, Chile and Argentina have reached an agreement in May 2014 to work towards removing roaming charges between these countries. Colombia has entered into an agreement with Peru as well with the same goal in mind. It is currently in the process of negotiating a regional agreement with Mexico, Peru and Chile. The goals are: promote competitiveness of the international mobile roaming market in the region (1), adopt measures to make users able to control their use of mobile services in the United States of America (2) and reduce roaming charges (3). As an example of Colombia's effort in reducing its international roaming prices, a decrease of 40% for voice services and 68% for mobile data are observed from 2013 to 2015 as shown in Fig. 2.5 [5].

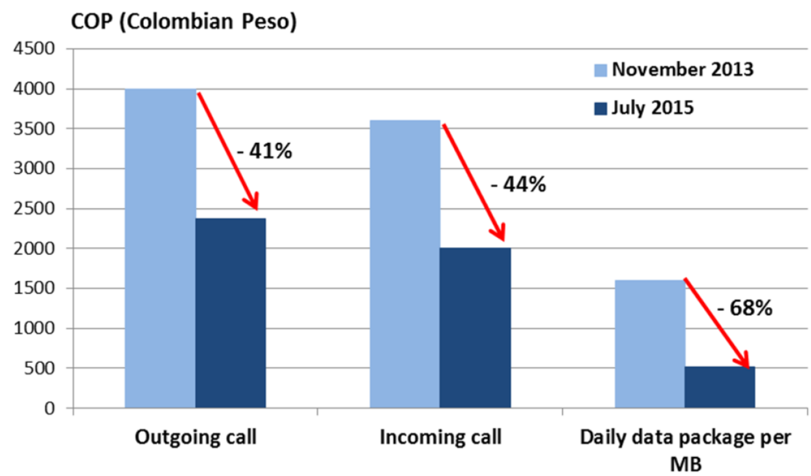

Fig. 2.5: Reduction of international roaming prices in Colombia between 2013 and 2015 [5]

In Central America, seven main operators introduced subscription plans which included roaming services back in 2015. The 'Without Border' initiative would allow end users in that region to only pay domestic rates set in their home country while roaming [6].

In 2018 Canada signed a commitment to end roaming charges by 2022 across North and South America at the Inter-American Telecommunication Commission (CITEL) meeting [12]. However, it is still unclear which concrete steps will be taken.

2.2.3 Africa

The Southern African Development Community (SADC) implemented a similar plan as the European Commission with the goal of evolving towards an RLAH-principle for its mobile users and a cost-based model for wholesale and retail price caps via three transition phases starting in 2014 and aiming at completion in 2020 [13].

In 2014, the East African Community (EAC) created the 'One Network Area' with a similar philosophy, i.e. to treat all mobile users in the community like the local customer. This is illustrated in the following aspects: (1) the customer can keep the same telephone number and SIM card across participating countries, (2) calls and text messages are charged at local rates, (3) prepaid customers are automatically charged in their home currency while post-paid customers are charged at local rates converted to their home currency upon billing [14].

The Economic Community of West African States (ECOWAS) has the intention to establish a single Information and Communications Technology (ICT) market in West Africa by: (1) harmonising the community regulation, (2) ensure total cross-border connectivity between all the ECOWAS countries, and (3) implementing a preferential community tariff to encourage and boost the use of roaming. The main objective is to also reduce the cost of roaming and potential termination rates. This process is currently still an ongoing process.

2.2.4 Asia

In December 2012 a bilateral agreement was reached between Israel and Russia to reduce roaming prices. It will use the EEA tariffs as "benchmark rates" for ongoing negotiations. Russia has also signed a bilateral agreements with Norwegian and Argentinian mobile operators with the aim of reducing roaming charges in the respective countries [5]. On the other hand, China signed a bilateral agreement with Denmark in order to reduce roaming charges between both countries [15].

Besides bilateral agreements, countries are also examining structural solutions to reduce roaming prices. For example, the Israeli Ministry of Communications announced a consultation procedure proposing that it would allow mobile operators, as well as other telecommunication service providers, to offer roaming as a separate service to end users of another mobile operator without the need to change the end user's phone number [16].

2.2.5 Australia

Australia and New Zealand agreed in 2013 to regulate mobile roaming prices through a bilateral agreement which gave regulators in both countries sufficient power to cooperate and intervene in the IMR market by, among others, allowing them to apply price caps on wholesale and retail roaming charges [17]. A legislative change is required for this agreement to be in effect and this is currently underway. The new legislation would allow the national regulatory bodies to impose retail and wholesale price caps on MNOs and require wholesale access obligations [5]. Australia and Japan have also begun discussions aimed at reducing roaming prices between both countries [18].

2.3 Effects of the roaming legislation

In the following subsections some important effects of the European roaming legislation will be discussed.

2.3.1 Decrease of wholesale and retail prices

The main purpose of all regulations and initiatives is to reduce the wholesale and retail prices together with the aim of creating sufficient competition in the IMR market. The EEA has made most progress so far. As mentioned before, since introducing the roaming legislation, the wholesale and retail prices have gradually decreased as a consequence of the implemented price caps. Although retail prices were generally more distant from the retail caps, wholesale prices remained close to the wholesale caps. Since 1 July 2017, retail surcharges of all three mobile services for were completely abolished. This means that end users, if they respect the fair use limits, are only paying their domestic rates for roaming in the EEA. At the time of writing, the wholesale caps for voice (€0.032 per minute) and text (€0,01 per SMS) are valid until 31 December 2019. These caps are currently being reviewed by the European Commission. The wholesale cap for data services is currently €4.50 per gigabyte. The current plan is to extend this cap to €3.50 on 1 January 2020. Then, to €3.00 on 1 January 2021 and finally to €2.50 per gigabyte on 1 January 2021.

In Asia, the wholesale and retail roaming rates between Singapore and Malaysia have been reduced by up to 30% for voice calls and up to 50% for SMS since 2011 [5]. Similar reductions appeared in other South East Asia nations.

The general trend of decreasing wholesale and retail prices is happening all around the world, however, the progression in each region is different. This is understandable since the collaborations and regulations did not start at the same time in all regions and each region has its own problems.

2.3.2 Increase in roaming volumes

As expected, the decrease in retail prices led to a large increase in roaming volumes, especially for data services. BEREC accumulated most data from its member states and published it in their quarterly international data benchmark reports in which the increasing trend of the roaming volumes can be observed. At the time of writing, data up to Q3-2018 has been made available by BEREC through these reports. Instead of focusing on the quarterly increases of the roaming data volumes (since the inception of RLAH in Q3-2017), it was decided to analyse the impact/increase of volumes over a whole year. Therefore, the average is taken over the percentage increase for roaming data use in each quarter after the inception of RLAH. This is shown in Fig. 2.6. The quarterly increase denotes the increase of a quarter before and after RLAH. For example, the increase in the first quarter after RLAH, is the increase in roaming data volumes between Q3-2017 and Q3-2016. This is then done for the following three quarters up till the most recent data of Q3-2018. Some countries are missing, because not all data of all four quarters was available. These countries are not included in the figure below.

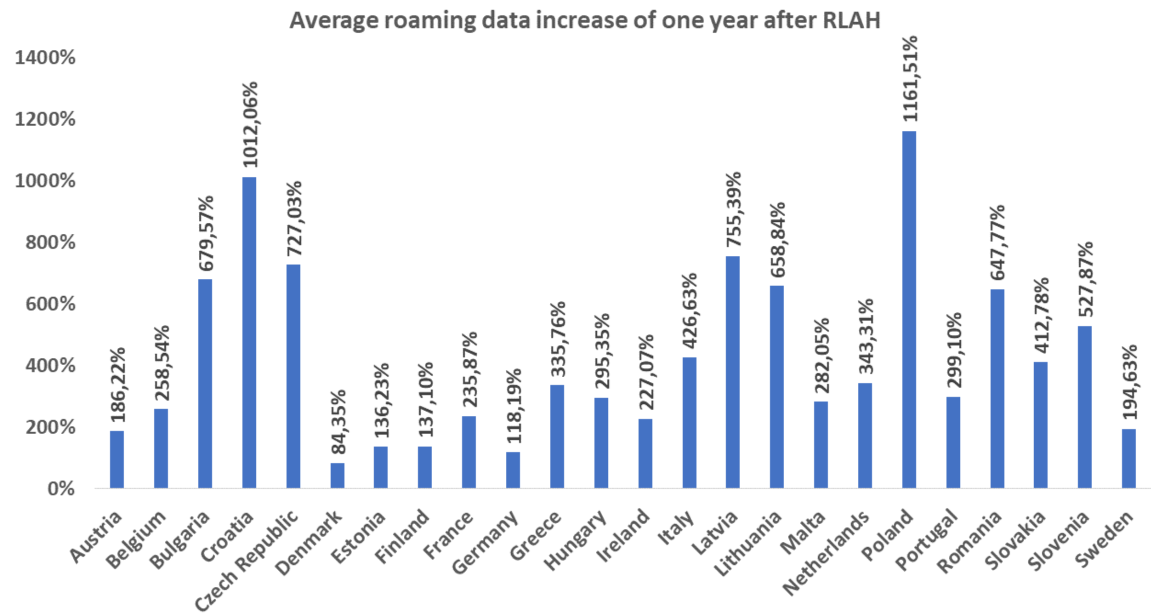

Fig. 2.6: Average percentage increase of roaming data use after RLAH (one year period)

The data in Fig. 2.6 is retrieved from the BEREC benchmark reports [10], [19]–[22]. It can be seen that experienced impact varies between countries. For example, why is the increase in Denmark so much smaller than in Poland? Denmark's roaming volumes belong to the higher ones in Europe so could it be that the users are already using more roaming data even before RLAH and is therefore, less affected by the change in legislation? These are the questions this dissertation aims to answer.

A similar scenario is to be expected in other parts of the world when RLAH is implemented for the local inhabitants. More data needs to be collected before further analysis can be made for each region specifically.

2.3.3 Impact of RLAH on different operators

With the inception of RLAH and the continued increase in roaming use, the experienced impact will be different for each type of operator. In the following paragraphs, these different potential impacts are discussed.

Loss of roaming revenues

From a business point of view, the first impact that all operators will experience is the loss of roaming revenue as a result of the abolishment of retail roaming surcharges. As an example, the retail roaming revenues for the main Belgian operators represented 8.2% of their total mobile turnover in 2015. This means that this part of the revenue is lost when RLAH became active. Other operators in the EEA will experience the same problem, the percentage revenue loss will of course vary depending on the roaming use of their subscribers.

Impact for MNOs: geographical location

The real impact of RLAH will be reflected in the changes in costs for operators. In the case of MNOs, the removal of roaming surcharges will be experienced differently depending on the country in which the operator is active. This is mainly due to the different travelling patterns of the end users. As a result, this will divide countries in two different categories: (1) net receiver, and (2) net sender [23]. A country with more incoming traffic compared to outgoing is called a net receiver while a country with more outgoing traffic compared to incoming is called a net sender. For net sender countries (e.g. Sweden), they have much more outgoing roaming traffic which makes the wholesale costs for these operators unbalanced (see the balanced/unbalanced pricing model explained in section 2.1.1). With the increase in roaming volumes, their wholesale costs will continue to increase while their retail roaming revenues are lost due to RLAH. On the other hand, net receiving countries (e.g. Spain) with more incoming traffic from visiting tourists have an incentive to keep the wholesale charges high as operators in these countries will most likely have to invest in network infrastructure upgrades to handle the traffic increases [24].

Impact of geographical coverage

Based on the area of coverage, operators can be dived into two categories (see section 2.1.2): (1) single country, and (2) cross-country. The latter will have a certain advantage compared to the former since its network spans across multiple countries. Cross-country operators will be able to obtain lower wholesale roaming prices by using their own network infrastructures [25]. The roaming traffic can be steered to their own network, in countries where they have infrastructure, which means that they do not have to pay additional wholesale prices for sending traffic compared to the single country operators [24]. Therefore, cross-country operators will have an advantage in terms of wholesale costs.

Impact for MVNOs

MVNOs resell capacity they rent from an MNO, they often take up only a small part of the domestic market. MVNOs incur costs when their customers are travelling, but they do not have wholesale incomes as they cannot host any roamers on their network[3]. They experience absolute traffic imbalances and often results is lack of bargaining power to negotiate wholesale roaming fees significantly below the wholesale caps [2], [26]. Therefore, these operators prefer that the EC keeps reducing the wholesale caps in order to mitigate an outflow of wholesale transaction which cannot be recovered on the retail level (due to the removal of retail roaming surcharges), and to assure a positive business case for these smaller players. If this issue is not tackled accordingly, RLAH might have a negative side effect on the level of competition within the national market [27]. In the latest BEREC opinion report to the EC regarding roaming [26], it was observed that MVNOs currently still lack negotiating power and therefore missing out on potential discounts on the wholesale price which are not pass on by the MNOs, the result is that the wholesale prices for MVNOs are still close to the current price caps imposed by the EC and higher than the wholesale prices paid between MNOs. One of the suggested guidelines was to force MNOs to pass on wholesale discounts to MVNOs. Another suggestion was to recalculate price caps in order to reflect the real cost of providing roaming since the wholesale prices between MNOs are significantly lower than the current caps.

2.4 Key issues in the international roaming market

In this section some of the current key issues in the international roaming market will be discussed.

[3] All types of MVNOs – except for Full MVNOs – are technically unable to accept any incoming roaming traffic. From a technical point of view, full MVNOs are able to accept incoming roaming traffic, though from an economic point of view this will never happen: the wholesale prices a full-MVNO could charge a DSP can never undercut the prices of its host-MNO. Taking into account even the slightest pricing margin, the wholesale prices a full-MVNO can offer to a DSP will always be higher than the ones from the underlying MNO; in other words, a DSP will always cooperate with the MNO for the simple reason that its wholesale prices are lower. [2]

2.4.1 Regulation vs. competition

Currently, the wholesale and retail roaming prices are regulated in the EEA. The International Telecommunication Union (ITU) [6] recognised that besides regulation and bilateral initiatives to reduce the roaming prices, structural measures must also be taken to encourage competition. Relying on excessive regulation in the market is not recommended. Currently, regulating bodies such as BEREC and NRAs are monitoring the markets in order to decide when it would be appropriated to adjust or remove existing regulations and price caps. The ideal landscape would be one with sufficient competition so that minimum regulation is needed to achieve cost-based wholesale and retail prices.

2.4.2 Wholesale and retail prices

The removal of retail roaming surcharges led to a decrease in revenue for the mobile operators. Together with price caps on wholesale prices, it is expected that operators will search for other ways to recover that lost revenue. One way to do this is to raise their domestic retail prices to compensate for their loss (the waterbed effect [28]) [2]. While the price reduction can be seen as a protection measure for end users, operators need to be protected as well. With the strict wholesale caps in the EEA, some smaller mobile operators risk becoming unsustainable, because they are unable to recover the basic costs of providing its customers with roaming in the first place. For this reason, the BEREC has provided some exceptions for these operators. They can ask for derogations in which they can charge a higher wholesale price than the existing price cap [21]. These derogations are mostly requested by net sending countries since their wholesale costs increases but these cannot be recuperated from its customers anymore. This shows that regulation alone is not enough, a long-term solution is required to provide a framework in which competition is encouraged.

2.4.3 Price transparency

According to a survey held by BEREC [29], 83% of the responding NRAs said that they received complaints on transparency issues in 2018. This is an increase compared to the previous period, where only 76% of the responding NRAs reported receiving such complaints. The survey data also shows that very few NRAs or consumer associations provided tariff comparisons. Only 21% of the responding NRAs reported that they provided updated information on their websites comparing tariffs that had a sustainability surcharge and 14% of them reported they provided updated information on their websites comparing roaming tariffs for non-EEA countries. Price transparency is important for end users and this lack of transparency is a potential reason why a portion of these users avoids roaming on their mobile devices. Mobile operators would actually benefit from being more transparent, because they would potentially attract more customers to use roaming which results in higher profits. Price transparency is thus beneficial for both parties and should be aimed for.

2.4.4 Balanced vs. unbalanced traffic

When an operator does not have inbound traffic, for example certain types of MVNOs, it has an disadvantage in negotiations with other operators. These (smaller) MVNOs risk paying a higher price for using another operator's network [1]. Mobile operators of net sending countries could potentially become unsustainable when they cannot recover the wholesale costs which they need to pay for allowing its own customers to roam. This is of course not always the case, but with the removal of roaming retail surcharges, this could become problematic for a number of operators. BEREC has acknowledged this problem and granted derogations to operators of certain sending countries, but it seems that not all of those operators are using this to surcharge their roaming customers. One of the reasons is that they are afraid of losing its customers to other competitors when they charge the extra amount since the landscape is becoming more competitive [26].

2.5 Current situation of the roaming market

In this section, some of BEREC's observations and opinions on the functioning of the roaming market, based on inputs from NRAs and operators (gathered via surveys), are described. This recent report [26] serves as guidelines for the European Commission on how to improve the current roaming regulation since the EC is set to publish a review of its legislation by the end of 2019.

2.5.1 Functioning of the retail market

Fair use policy (FUP) and permanent roaming

The survey reveals that the vast majority of the operators apply a FUP: 95% of the MNOs and 78% of the MVNOs responding to the survey. The operators largely notified their FUP to the NRA (91% MNOs and 86% of the MVNOs) and most of them also implemented a simple and transparent procedure for their customers to address complaints. However, six MNOs and three MVNOs reported not to have provided their customers with such a mechanism to receive complaints. Nonetheless, it seems that the FUP is useful since a vast majority is applying them. However, the EC should try to improve their policy to support more MVNOs in using a FUP.

The survey also shows that operators generally comply with the legal provisions when applying a FUP. However, taking into account the input provided, BEREC sees the need to clarify some rules that relate to the application of a FUP, because the provisions are quite complex to handle when it comes to assessing a FUP notified by operators. BEREC considers that the EC could clarify the rules applicable when the formula for calculating a FUP for open data bundles yields a higher roaming allowance than the domestic allowance. For such cases, BEREC recommends the rule to set the roaming allowance to be equal to the domestic allowance.

The possibility to include conditions against permanent roaming or abusive use of wholesale roaming access in their reference offer is only partly used (only 10% of the operators). Therefore, it can be assumed that most operators see no necessity to include specific conditions to prevent permanent or anomalous or abusive usage of wholesale roaming access in their reference offer. Although the possibility to include such clauses is not widely used, it is relevant to highlight that operators who have detected businesses based on permanent roaming have decided to pursue commercial agreements with access seekers rather than deny access, which would have led to a dispute with the relevant NRA. In summary, BEREC considers that the Roaming Regulation is currently sufficient in preventing permanent roaming or anomalous or abusive usage on wholesale level with the measures provided.

The BEREC analysis shows that the introduction of RLAH had no major impact on prices or consumption patterns for both domestic and RoW services. Furthermore, there is currently no indication that RLAH has any serious impact on the availability of domestic offers, which is further corroborated by the evidence available to BEREC that the overall domestic tariff structure remains in most cases unchanged. However, BEREC notes that there are some changes to domestic tariff plans (some of which were observed before RLAH was introduced, anticipating the changes that would occur) and a high share of subscribers with domestic-only tariffs was reported by some countries. This is mainly due to operators offering cheap tariff plans without the possibility to roam. This is understandable because this will allow operators to reduce the wholesale costs they have to pay and also attract more end users who only use mobile services domestically.

2.5.2 Functioning of the wholesale market

Situation for MVNOs

According to BEREC, MVNOs lack of a radio network to offer connectivity to inbound roamers, and in general limited resources for managing direct wholesale roaming, makes most of them dependent on some form of resale access. Lack of negotiation power due to size and, for some of them, the dependency on the host, makes it challenging to achieve discounts or better rates than the regulated caps. This group of operators furthermore have no wholesale roaming revenues to balance the wholesale cost, which makes their situation challenging and very different from MNOs (as explained in section 2.3.3). Therefore, BEREC suggests the following possible measures that the European Commission could introduce in any update of the Roaming provisions in order to increase the competitive strength for MVNOs:

- Reducing wholesale caps, taking into account that MNOs need to recover their efficiently incurred costs to provide wholesale roaming services. This is considered an efficient and transparent measure.

- Obliging the host MNOs to pass the discounts they get for wholesale roaming services on to the MVNOs. Although this measure would ensure equal terms for competition between MVNOs and MNOs, BEREC considers that this measure is very complex to implement and would require the definition of a monitoring process by NRAs.

Additionally, there is a need to clarify that regulated maximum caps also applies for wholesale resale access to MVNOs. The following measures are proposed to the EC:

- Make sure that wholesale caps also apply to alternative wholesale roaming solutions like sponsored roaming. However, this does not prevent providers of such wholesale solutions from charging additionally for other services they offer.

- Include measures for incoming roaming calls for MVNOs.

2.5.3 Impact of RLAH on quality of service (QoS)

Some end users raised complaints about lower speeds while roaming in the EEA. Certain cases were reported where operators restrict speed or technologies (only 3G available despite vast majority of operators in the EU offer 4G services) when customers are travelling abroad. One of the reasons to reduce the speed or restrict the quality is to reduce data traffic (and thereby wholesale costs). The intention of the Roaming Regulation is to allow roaming customers to use the service like at home. Although the Roaming Regulation does not provide any obligations in terms of QoS requirements, domestic operators should not purposely lower the QoS than the one offered at home according to BEREC. In addition, operators should be transparent towards the customers in terms of QoS in a roaming situation (e.g. website, contracts, etc.). Therefore, BEREC suggests that the EC could further investigate imposing more specific obligations for the home network operator in this regard in any potential update of the Roaming Regulation. Also, the Roaming Regulation could be updated/improved in order to enable all operators, in particular small operators, to offer the same QoS for roaming as domestically in a sustainable way.

2.5.4 Functioning of the derogation mechanism

BEREC assessed the cases where MNOs and MVNOs requested a derogation from RLAH and analysed whether the mechanism provided by the regulation was working effectively to achieve the sustainability of the roaming market.

In June 2017, a total of 30 derogations had been granted by NRAs. For the period 15 June 2017 to 14 June 2018, 17 NRAs received applications for sustainability surcharges. In total 57 applications were received, 46 of which were granted and 11 of which were refused in this period. For the period 15 June 2018 to 14 June 2019, only 10 NRAs received applications for derogation. In total 37 applications were received, all of which were granted.

Overall, this means that only a small percentage of operators active on the market made use of the derogation mechanism. Moreover, these derogations concern essentially smaller MNOs and MVNOs which have a very low share of a given market. Four countries (Poland, Finland, Estonia, and Lithuania) are an exception in that matter according to BEREC, as derogations were granted to major operators due to very low retail prices and high wholesale traffic asymmetry compared with other member states.

In BEREC's observations, MVNOs are the primary users of the sustainability derogation given their specific situation on the wholesale roaming market. Over the first year of RLAH, about two-third of the derogations have been granted to MVNOs (30 versus 14). Table 2.1 summarises the number of derogations in place as of Q1-2019 per country, split by type of operator (MNO and MVNO). It can be seen that nearly two-third of the derogations concern MVNOs. Regarding the use of the derogation granted, in a majority of the cases, the derogation was not applied to all tariff plans.

Table 2.1: Sustainability derogation in place as of Q1-2019 [26]

Country	MNO	MVNO	Market share (%)
Austria	0	2	≈ 1
Belgium	0	1	< 5
Estonia	3	0	100
Finland	3	1	100
France	0	6	3
Italy	0	3	2.7
Lithuania	3	1	≈ 100
Poland	4	8	100
Romania	1	0	13
Slovenia	0	1	2.6

Overall, this indicates that operators granted a derogation still strive to apply RLAH as far as possible in order to remain competitive in the market. Some NRAs have already started to assess derogation renewals for the next yearly period. It is expected that the number of derogations for the next periods will decrease: approximately 30 derogations to be granted in 2019 and approximately 20 derogations in 2020.

In conclusion, the derogation mechanism concerns a small and shrinking part of the market in most member states. However, this mechanism remains a tool for some operators to achieve the overall sustainability of RLAH (especially for MVNOs and operators from markets with low retail ARPU (average revenue per user) and markets with a high level of unbalanced traffic). Taking into account the feedback from NRAs, BEREC suggests the recalculation of wholesale caps in order to reflect the actual costs more accurately as well as taking the costs of balanced traffic into account. In this way, the difference in wholesale costs paid by MNOs and MVNOs will decrease. Another noticeable trend is the strong competitive pressure on operators, which prevents them from extensively applying the surcharges even after they have obtained the derogation.

2.6 Insights and research questions

From analysing the roaming evolution in the international roaming markets, it can be observed that there is still a long way to go before there is sufficient competition in the market so that a minimum of regulation is required. Currently, the European roaming legislation seems to have progressed the furthest as a result of the continuous efforts of the EC to revise the regulations and strategies to encourage competitiveness between operators as well as protecting both consumer and operator. Other regions and countries have taken the European roaming legislation as benchmark for developing their own roaming legislation. However, the European roaming legislation is definitely not completed since it is still being reviewed in order to adapt to the evolving market.

It was observed that the increase in roaming volumes, especially for data services, was very large due the RLAH. This proposes new questions and challenges for mobile operators. On the one hand, operators of receiving countries have to evaluate whether their network infrastructure will be able to handle the increase in inbound roaming traffic. If not, an infrastructure upgrade is required which results in additional costs. On the other hand, operators of sending countries may face the challenge of a large increase in inter-operator tariffs they have to pay when their customers start to use more data services in EEA countries. With the disappearance of their retail roaming revenues due to RLAH, European operators have to re-evaluate their strategy in order to remain competitive and recover the lost revenues. The first step towards finding a solution for this problem is to analyse how the roaming volumes are expected to increase in the near-future. From there, operators can assess the real impact for them by using the roaming volumes predictions as input in their cost model. The results of this cost model will allow operators to estimate the change in total costs which is the real impact of RLAH. This leads to the research questions of this master's dissertation: Which factors affect the usage of mobile roaming services (1) and what is the impact of RLAH (2) on different countries/regions? To answer these questions, a conceptual model needs to be developed which contains influencing factors on the roaming volumes. Some factors will be region specific. From the forecasted volumes, the impact for each country can be analysed. In the following chapter, existing studies or models related to predicting mobile internet use will be analysed from which the potential influencing factors will be retrieved.

Chapter 3

Literature review

In this chapter, a literature review of existing studies on the use of mobile services is performed in order to analyse and determine possible factors influencing the mobile data use. The insights gained through this literature review serve as inputs for the conceptual model.

As mentioned in Chapter 2, there are three mobile services: voice, text, and data. Although an increase in use for all three services is expected, the focus of the model in this dissertation will be on the roaming data use. The reason for this choice is because data volumes contribute a lot more in terms of load on an operator's network compared to voice or text services. Together with the digitalisation in our civilisation and increased smartphone and mobile applications use, which is gradually replacing the need to use text or voice services with applications like social media and over-the-top services (e.g. Whatsapp), it seems reasonable to focus on the analysis of data use instead of the other services.

Before the data collection process, it is useful to first analyse which factors are useful. This is done through analysing publications found via Google Scholar in which the following keywords were used: mobile use, mobile data use, mobile services use, factors influencing mobile use. The EC's publications related to roaming were used as well. The contributions found through this method were fragmented in the sense that the papers studied a wide variety of topics or applications related to mobile internet use, e.g. mobile internet use related to social media [30] or the use of mobile banking [31]. Few studies, that were found via this method, focused on forecasting roaming volumes specifically. Cisco does publish its Visual Networking Index (VNI) forecasts for global fixed and mobile internet traffic, e.g. [32]. However, these focus on what the actual forecasted numbers are without going into detail which methods they used to obtain the forecasts.

Gerpott and Thomas [33] observed that empirical findings on mobile internet use are widely spread across many disciplines such as communications engineering, computer science, electronic business, information systems, management, marketing, social science and telecommunications. This means that the existing work is very fragmented and scattered across a wide variety of disciplines. Furthermore, these studies vary in quality with regard to sampling procedures as well as reliability and validity of variable measurement. For this reason, it is difficult to compare conclusions across different studies. Gerpott and Thomas [33] attempted comparing results of different studies in which the same factor was analysed by using a meta-analysis. This is a statistical procedure for combining data from multiple studies from which better conclusions can be drawn. When the treatment effect (or effect size) is (1) consistent or (2) varies from one study to the next, a meta-analysis can be used to identify either (1) this common effect or (2) the reason for variation [34]. Decisions about the utility of an intervention or the validity of a hypothesis should not be based on results of a single study, because these results typically vary from one study to the next. Therefore, a mechanism is needed to synthesise data across studies. Narrative reviews had been used for this purpose, but this type of review is largely subjective (different experts can come to different conclusions) and it becomes more complex when there are more than a few studies involved. A meta-analysis, by contrast, applies objective formulas (much as one would apply statistics to data within a single study), and can be used with any number of studies [34]. Therefore, a comparison of significant factors found across existing studies is an important first step in determining which factors have a potential influence on mobile data and roaming use.

3.1 Comparing results of different studies

Gerpott and Thomas [33] performed a meta-analysis on the literature dealing with empirical research on mobile internet usage. They reviewed 175 scholarly empirical publications on mobile internet (MI) usage intensity levels and potential influencing factors of respective usage behaviour at individual subscriber level. These include 80 journal articles, 86 conference papers, three dissertations, four book chapters and two working papers. For the full list of these contributions, the appendix section of Gerpott and Thomas [33] can be consulted. They observed that the inter-individual variance in MI usage was very large. This raises the question which factors contribute to explaining MI usage differences. They performed a qualitative review and a meta-analysis of correlations between 22 variables grouped into four categories on one side and MI usage criteria on the other. In the following subsections, the useful findings of Gerpott and Thomas [33], in the context of this dissertation, will be summarised. The procedures followed by Gerpott and Thomas [33] for the meta-analysis were based on the following works: Hunter and Schmidt [35] and Peterson and Brown [36].

3.2 Mobile internet usage measurement approach

[33] found that studies take very different approaches in capturing the MI usage intensity, each with specific strengths and weaknesses. They have mentioned that MI usage intensity can be measured using two methods: subjective or objective. These methods are explained in the following subsections.

3.2.1 Subjective methods

Subjective methods are self-estimates of personal usage amounts. These are often collected in a one-shot survey or, infrequently, via a longitudinal panel. Over half of the analysed contributions in [33] use surveys to collect data with most of them only relying on self-reported usage. One of the possible problems is that self-report measures are likely to overestimate the actual MI usage since MI use is socially potentially regarded as an indicator of a "progressive active lifestyle". Correspondingly, respondents over-report their MI usage in order to improve the image they suspect to convey to the investigator [37]. Another problem could be that self-reports require respondents to remember detailed facets of past MI usage. Most of the analysed survey-based studies do not address the cognitive burden of MI users when trying to recall behaviours by introducing diaries for recording personal MI usage because collection and analysis of diary entries are arduous. Also, the willingness to participate in diary keeping is probably low and biased towards customers sharing a strong interest in MI usage research. Since investigators mostly do not support MI end users with tools to recall and estimate the duration, frequency, data volumes or the number covered different services/applications, subjective MI usage measurements are likely to be limited in accuracy and level granularity [38]. However, surveys are the only approach providing the opportunity to include explanatory variables which are not readily observable, e.g. user motives, service perceptions and satisfaction [39]. This is one of the main reasons why a large number of researchers keep using surveys for data collection despite the potential problems of participants over-reporting their MI usage.

3.2.2 Objective methods

Objective methods rely on "system-captured" MI usage metrics [40], e.g. using an operator's network to collect end users' data usage. Depending on the measurement point in the system, these methods can be classified into three groups: (1) handset monitoring, (2) traffic measurement, and (3) usage billing according to Kivi [39] and Smura, Kivi, and Töyli [41]. Each of these methods has its own strengths and weaknesses.

With handset monitoring, a software must be installed in the end user's device(s) to track the MI usage, e.g. by in-app tracking. While this provides accurate measurements for each end user, potential problems may arise from using these kinds of monitoring software. First, the software is programmed for a particular handset operating system (OS) and is thus not directly compatible across other operating systems. One could argue that this problem is limited since most smartphones or tablets nowadays are either Android (Google) or iOS (Apple) based. Secondly, comprehensive tracking software is likely to cause privacy concerns [38]. As a result, it is mostly used within closed customer groups with a small number of participants. Subjective explanatory variables such as assessments of previous MI service encounters cannot be obtained via this method. To overcome this problem, surveys can be taken from the small number of participants since researchers are usually in close contact with them [33]. Mobile operators could potentially gather more data since they have a large subscription base which could participate in the survey.

Mobile operators can collect MI usage data of their customers directly from the network infrastructure via traffic measurement. These can differentiate down- and uploads as well as particular, but not all service classes (e.g. Voice over IP). However, individual users or specific user interactions can hardly be differentiated from each other without complementary measurements. Also, the separation of user clusters is limited to few characteristics, e.g. device category (mobile handsets, laptops or tablets) or operating system [33].

The third option is to analyse the mobile operator's billing systems, which register the demand for chargeable mobile data services. Normally, billing systems provide monthly data volume as a usage measure, which does not allow differentiating services or service classes. Billing data can be matched with supplementary information derived from the contracting party's master data file, but one cannot be sure that the actual user is always identical with the individual recorded as contract holder [33].

3.2.3 Comparison of both methods

From a scientific point of view, the objective methods would result in the most accurate measurements of ones MI usage. However, in order to apply these methods, one must be able to access the technical installations of mobile operators or device manufactures. However, operators are often reluctant to grant access to researchers, which is understandable since they have to protect their own interests as well as the privacy of their customers. Consequently, about 40% of the analysed studies completely refrained from collecting data using objective methods and purely focused on self-reported data collected through surveys [33]. On the other hand, not all influencing factors can be derived from objective data alone as mentioned before. Therefore, self-reported survey data is still valuable and should be combined with objectively collected data in order to analyse a wide-range of possible factors.

Countries have different characteristics; therefore, one could argue that comparing these studies cannot distinguish the differences on country level. However, due to the complexity of predicting the roaming volumes, it is unlikely that one model will be able to identify all country level differences. Therefore, it is suggested to start from a broader conceptual model containing more factors and gradually eliminate the insignificant ones for different countries/regions. Just because one factor does not seem to have an impact in one country does not mean that the same goes for another. Therefore, a specific model should be developed for a region or a group of countries with each its set of relevant factors. The country/region specific differences are then expressed through the different factors and weights used in the different models. This approach requires much more data samples due to the larger number of factors included in the model.

Another important factor to take into account is that most studies are cross-sectional meaning that each study is done at a specific point of time. When these studies are compared to each other, researchers often do not take into account a population's evolution over time. For example, a factor that was not seen as important in last year's study may be important in this year's study. For this reason, researchers have suggested to perform longitudinal studies in order to get a better understanding on the evolution of these factors.

3.3 Theoretical frameworks explored by existing studies

Currently, there is no single theoretical framework for selecting influencing factors on the mobile or roaming data use specifically. A considerable share of the analysed studies are "explorative" in the sense that they proceed without a specific conceptual framework. Publications in this category mainly look at personal characteristics of a mobile internet user, its country of residence or factual use conditions as potential factors [33].

In theory-based papers, a broad range of conceptual frameworks are used to justify the selection of potential factors. The most prominent conceptual models in this context are: Technology Acceptance Model (TAM) of Davis, Bagozzi, and Warshaw [42], Theory of Planned Behaviour (TPB) of Ajzen [43], and Innovation Diffusion Theory of Rogers [44].

The main idea of the original TAM is that perceived usefulness (PU) and the perceived ease of use (PEOU) affect one's MI usage indirectly through attitude towards using MI as well as behavioural intention to use it. PU and PEOU are influenced by some external variables. When applying this model, the researcher can choose which external variables he or she wants to include in the model. These variables are then tested statistically to determine if they are significant. The general TAM model is shown in Fig. 3.1. Due to its generality, it can be adapted for a wide variety of technology adoptions in different fields. For example, analysing online consumer behaviour [45] or examining physicians' acceptance of telemedicine technology [46]. Although King and He [47] confirmed the validity and robustness of TAM through analysing and comparing 88 TAM-related papers, Chuttur [48] identified some important limitations of this model, e.g. the use self-reported data volumes is used instead of objectively measured ones. This is, according to some researchers, inaccurate and prone to errors as explained earlier in section 3.2.1. Yousafzai, Foxall, and Pallister [49] also noticed that, in contrast to the large number of studies applying TAM to explain and predict the voluntary use of systems, very few studies considered systems that were for mandatory use. However, in real life settings, most organisations usually require users to use the system available with little choice for alternatives [50]. Due to these limitations, the TAM has been criticised leading the original proposers to attempt redefining it several times.

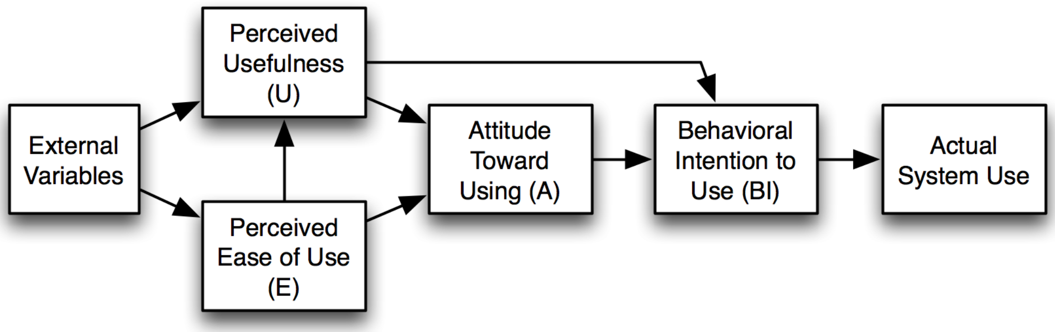

Fig. 3.1: General Technology Acceptance Model (TAM) [42]

The TPB (Fig. 3.2) hypothesizes that the effects of three constructs (attitude, subjective norm, perceived behavioural control/self-efficacy) on MI usage are mediated by behavioural intentions [43]. Despite its frequent use in health and science related studies involving human behaviour, the main criticism is that this model cannot explain sufficient variability in all volitional behaviour with only four explanatory variables [51]. Sniehotta [52] also argued that one's emotions at the time of interviewing are ignored despite being relevant to the model as these emotions can influence beliefs and other constructs of the model.

According to Rogers [44]'s Innovation Diffusion Theory (IDT), mobile internet offerings can be classified as an innovation. Consequently, perceptions of five innovation attributes (relative advantage, compatibility, complexity, trial-ability, and observability) should be considered as key factors in explaining initial MI adoption and subsequent MI usage behaviours. Two of these attributes resemble TAM constructs (usefulness and relative advantage; ease of use and complexity), therefore, Wu and Wang [53] proposed that the constructs employed in TAM are fundamentally a subset of the perceived innovation characteristics. Fig. 3.3 shows how Rogers [44] interpreted the adoption of technology in general. The blue curve represents how the population adopts a new technology which is assumed to be normally distributed. The orange curve, sometimes called the "S-curve", shows the adoption rate of the technology. It increases slowly in the beginning due to the small amount of people that is aware of this new

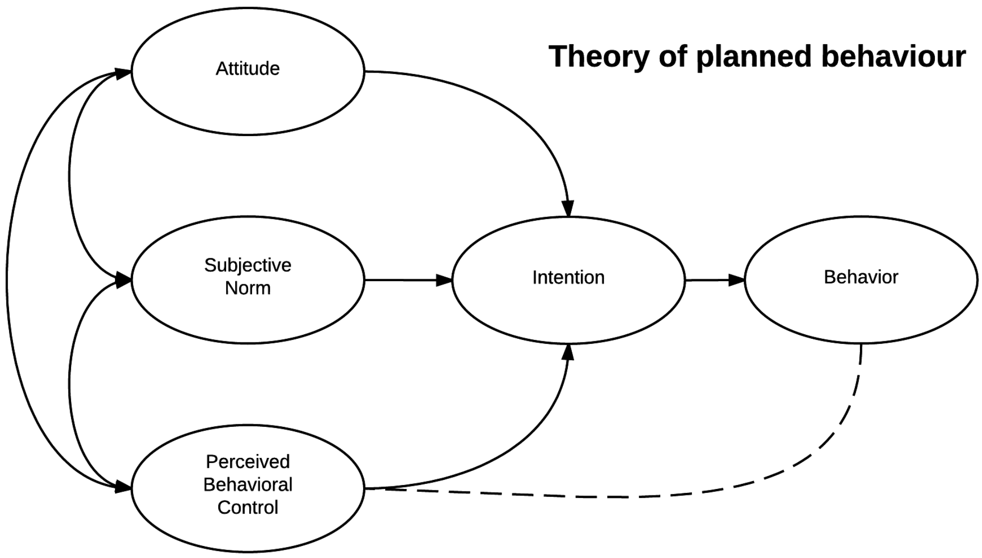

Fig. 3.2: Theory of Planned Behaviour (TPB) Model [43]

technology's existence. By the time more people know about it, the Early-adopters start using it as well leading to a faster increase. The inflection point is situated on the location where the latter half of the population starts adopting this technology. By the time the Laggards start adopting it, the increase is slowed down, the final point in this model is when the whole population uses this technology resulting in an adoption rate of 100%.

All in all, TAM-, TPB- and IDT-based investigations tend to focus on the stated behavioural usage intention and perceptions of MI attributes [33]. Since these three models were frequently used in past studies, researchers have attempted to unify the existing models into one larger model with the aim of eliminating the limitations described above. This resulted in Venkatesh, Morris, Davis, *et al.* [54]'s Unified Theory of Acceptance and Use of Technology (UTAUT) model. This theory holds that there are four key constructs: (1) performance expectancy, (2) effort expectancy, (3) social influence, and (4) facilitating conditions. The first three constructs are direct determinants of usage intention and behaviour while the fourth one is a direct determinant of user behaviour. Gender, Age, Experience, and Voluntariness of use are posited to moderate the impact of four key constructs on usage intention and behaviour. This theoretical model was developed through a review and consolidation of the constructs in eight models that earlier research had employed to explain information systems usage behaviour. These are: (1) theory of reasoned action, (2) technology acceptance model, (3) motivational model, (4) theory of planned behaviour, (5) a combined theory of planned behaviour and technology acceptance model, (6) model of personal computer use, (7) diffusion of innovations theory, and (8) social cognitive theory. A detailed explanation of each model can be found in

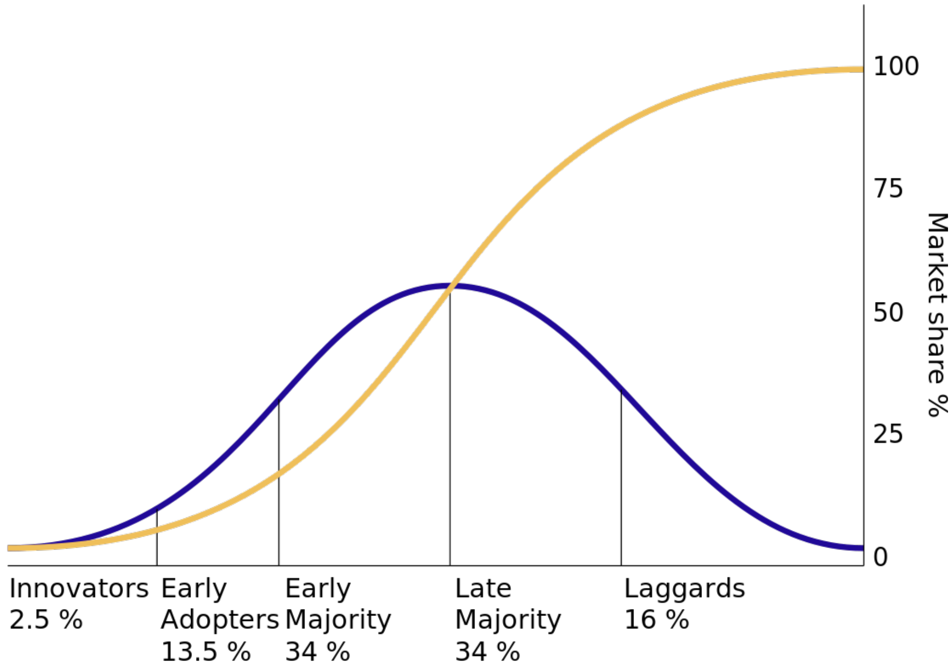

Fig. 3.3: Innovation Diffusion Theory (IDT) [44]

Venkatesh, Morris, Davis, *et al.* [54]. The authors of UTAUT also validated their model and found that in a longitudinal study, the model accounted for 70% of the variance in behavioural intention to use and about 50% in actual use. This theory now contains more factors in order to explain more variability of the studied phenomenon and has been frequently used since its formulation. For example, Verhoeven, Heerwegh, and De Wit [55] applied this model to study the computer use frequency of university students and found that UTAUT was also useful in explaining the varying frequencies of computer use and differences in ICT skills in secondary school compared to the university. Despite the efforts of Venkatesh, Morris, Davis, *et al.* [54] to incorporate more factors in this framework, it has also received its fair share of criticisms. Bagozzi [56] expressed his critique of the model, and its subsequent extensions, by stating: "UTAUT is a well-meaning and thoughtful presentation," but that it presents a model with 41 independent variables for predicting intentions and at least eight independent variables for predicting behaviour." He also claimed that it contributed to the study of technology adoption "reaching a stage of chaos." He proposed a unified theory that coheres the "many splinters of knowledge" to explain decision making instead.

It is clear from the observations above that no model will be perfect for analysing all possible technology acceptance and usage related topics. Researchers draw elements from different models to create their own model to investigate the adoption or drivers of a specific technology. This should also be done in the context of this dissertation. The literature can only provide an overview of which factors were found to be useful in explaining technology acceptance/usage in general. Whether they will actually have an impact in the studied phenomenon remains to be determined from testing the factors using statistical procedures.

3.4 Factors analysed in existing studies

Due to the absence of a single theoretical framework for selecting influencing factors of mobile internet usage and the broad range of antecedents covered in extant work, Gerpott and Thomas [33] proposed to structure the potential factors into four categories: (1) country, (2) personal user characteristics, (3) perceptions of MI attributes, and (4) factual MI usage conditions. They summarised all potential factors of the analysed studies and performed a meta-analysis on the correlations between individual factors and MI usage intention. In the following subsections, the results of their analysis will be discussed. From these results, the most promising factors will be selected for the conceptual model.

3.4.1 Country

In sixteen of the analysed studies, samples from at least two countries were compared. From these cross-country comparisons, all studies, except for Nielsen and Fjuk [57], reported significant deviations in various mobile internet usage indicator levels ([58]–[72]).

According to Gerpott and Thomas [33], it is difficult to derive the precise meaning of the detected country differences, because studies vary significantly in respect to the representativity of the samples used. They also observed that the term "country" is a catch-all variable which reflects a plethora of diverging background features, such as political and cultural values, economic development stages or technical capabilities of telecommunications infrastructures in place [73]–[75]. This is understandable since the mobile penetration is also different in each country. The mobile penetration differences in European countries may be smaller, but between developing and developed countries in terms of technology acceptance, the difference in studied topic (in this case the MI usage), will probably be significantly different.

3.4.2 Personal characteristics

Personal characteristics encompass socio-demographic variables, presumed MI valuation by close contacts, mobile service-related self-efficacy, attitudes on new offerings in general, various facets of MI experience and usage of established mobile communication services. The characteristics that were included in the analysed studies are shown in Table 3.1.

According to the analysis of Gerpott and Thomas [33], the most correlated factors are educational level and self-assessed general openness for innovative technical durables (innovativeness). In the analysed studies, each factor was tested individually for correlation with the MI

Table 3.1: Studied personal characteristics on MI usage [33]

Personal characteristic	K	n
Age	22	27027
Gender	22	28358
Income	10	5788
Educational level	7	5757
Subjective norm	7	3154
Technology self-efficacy	3	2069
Innovativeness	3	1479
MNO tenure	3	18921
Extent of MI experience	11	25556

K = number of studies which include the characteristic
n = number of study objects

usage intensity. One factor is of course not enough to explain all the variability. Therefore, it is worthwhile to search for other factors which may account for a considerable amount of variation in the correlations between personal characteristics and MI usage intensity. Gerpott and Thomas [33] also observed that it was not always clear whether the correlation is positive or negative. This suggests that the correlation is dependent on the studied population. This will most likely also be the case in European countries.

O'Doherty, Hill, Mackay, *et al.* [76] observed that the type of services used also co-varies with the educational level of the user in the sense that a hedonic MI use is more prevalent in less educated than in higher educated customer groups.

According to Gerpott and Thomas [33], existing studies suggested that mobile voice and SMS on the one hand and MI access on the other to date have been *complementary* and *not substitutive* modes of telecommunication. However, since some of the analysed studies dates back to over a decade ago and with the increasing use of MI-based voice and messaging apps on smartphones, it would be useful to test whether this conclusion still holds in the current situation [33].

To summarise, the potential factors analysed in this category displayed a small to medium effect on the MI usage intensity according to Lee, Kim, Choi, *et al.* [74]. Additionally, it is worthwhile to search for other variables which may account for the considerable variance of the correlations between personal characteristics and MI usage intensity.

3.4.3 Behaviour intention and attribute perceptions

The third category of potential factors on MI usage contains the behavioural MI use intention and perceptions of six MI attributes which researchers have drawn from the three different theoretical frameworks discussed in section 3.3. The results of the analysed factors by Gerpott and Thomas [33] are summarised in Table 3.2.

Table 3.2: Studied behaviour intention and attribute perceptions on MI usage [33]

MI intention/perception	K	n
Intention to use MI	11	3693
Usefulness	11	5961
Ease of use	7	3705
Monetary value	6	3452
Enjoyment	6	4390
Facilitating conditions	3	2185
Customer satisfaction	4	1808

K = number of studies which include the characteristic
n = number of study objects

Gerpott and Thomas [33] found that all factors in this category have a positive influence on the MI usage across all analysed studies. The least correlated factors were found to be customer satisfaction and facilitating conditions. Facilitating conditions refer to a person's beliefs that her MNO provides those resources (e.g. network bandwidth, customer care) which facilitate the use of mobile internet [54]. The largest correlation was observed for enjoyment, i.e. the hedonic or intrinsic value experienced when accessing the internet over a cellular data network. Furthermore, the behavioural intention to use MI and ease of use perceptions also showed positive associations with MI usage [33].

In contrast, Gerpott and Thomas [33] observed that the analysed studies were more ambiguous for the four perceptual constructs: (1) MI usefulness, (2) monetary value of MI, (3) facilitating conditions, and (4) customer satisfaction ratings. They claimed that the results leave it open whether the correlation between one of these factors and MI usage is positive or negative. They noted that this finding was interesting from a theoretical perspective, because it runs counter to the TAM and TPB frameworks which both posit that MI attribute assessments affect MI usage purely indirectly through MI usage intentions.

To summarise, MI-related behaviour intention and perceived attributes of MNO customers tended to display larger mean effects on MI usage than the personal characteristics factors according to Gerpott and Thomas [33]. They also suggested that further investigation is needed for factors in this category to determine potential hidden variables which may account for a considerable amount of variation between the behavioural intention factors and the MI usage.

3.4.4 Factual use conditions

The last category of potential factors is the "factual use conditions". These conditions act as positive or negative incentives for a person to resort to mobile internet functions. In the work on user interface designs, these were also characterised as "affordances". This category encompasses objective technical performance parameters of the network (e.g. transmission speed and latency) of a user's mobile service provider and of her access device (e.g. screen size and resolution, memory capacity). Additionally, it covers a customer's mobile internet tariff type as an important objective commercial condition of MI usage. Gerpott and Thomas [33] claimed that although there were quite a number of studies which contained some observations concerning the impacts of various factual use conditions, the large majority of them failed to give the complete information necessary to conduct meta-analytic effect size calculations. Therefore, a quantitative analysis was only possible for the three factors shown in Table 3.3.

Table 3.3: Studied factual use conditions on MI usage [33]

Factual MI use condition	K	n
Type of access device	21	1268
Network and device capability	10	4866
Tariff type	6	1047

K = number of studies which include the characteristic
n = number of study objects

The first factor is the type of MI access device one uses to access mobile internet. In general, researchers (e.g. Gerpott, Thomas, and Weichert [77]) have grouped customer MI premises equipment roughly into two categories: computer-centric appliances (laptops, netbooks and tablets) and MI-enabled (smart)phones. Available evidence indicates that the average mobile internet data use of people with computer-centric devices is significantly larger than that of their counterparts equipped with smartphones ([78]–[83]). This is reflected in Gerpott and Thomas [33]'s meta-analytic findings of three studies which reported sufficient data to derive effect sizes. They found that the type of access device (1 = computer-centric, 0 = other) had a medium effect on the MI usage and was positively correlated to it.

For the device type factor, supplementary insights can be obtained from Jin, Duffield, Gerber, *et al.* [84]. They detected that the proportion of “heavy” users, who generated much more data than the average user, was significantly higher among persons equipped with computer-centric devices compared to smartphone customers. Furthermore, several studies ([85]–[88]) showed that laptops were used less frequently and during a shorter period of time per day than smartphones. He, Lee, Pan, *et al.* [79] and Liu, Chuah, Zang, *et al.* [81] also revealed that the proportion of peer-to-peer, (video) streaming, and e-commerce related traffic is higher for computer-centric devices compared to smartphones. Thus, extant evidence points in the direction of laptop customers being the more "data hungry" MI types compared to smartphone users which is intuitively easy to understand. However, partly due to the low proportion of studies reporting effect sizes, it is not completely clear from this study which service types are responsible for the differences in usage levels between people relying on a laptop versus a smartphone when accessing the internet via radio networks [33].

The second factor is network and device capability in a technical sense. From the end user's perspective, radio network coverage and transmission performance of successive generations of mobile data infrastructures can hardly be isolated from the device capabilities. Because network and device are reciprocally linked in shaping the quality of an end user's experiences when accessing the internet over a cellular network. According to Gerpott and Thomas [33], further empirical pieces for which it was not possible to quantify pertinent effect sizes indicate that:

1. The MI usage of MNO customers, who had access to radio network generations enabling higher transmission speed, significantly exceeded that of their counterparts who only had the option to access an earlier, less powerful network generation [89], [90].
2. For the subgroup of people with smartphones, the technical capability level of their device exerted a significantly positive impact on MI usage [65], [81], [82], [91]–[98].

A third non-technical, but commercial factual MI use condition, is the tariff type of a customer. This variable indicates whether an individual selected a 'flat' MI rate plan whose charge is independent of one's MI usage volume (mostly data volume generated) or whether it is in some variant of usage-dependent pricing schemes. The meta-analysis showed a positive correlation between this factor and the MI usage intention. This suggests that mobile internet tariff type has a considerable impact on MI usage levels.

3.4.5 Moderator analysis

In the final step of Gerpott and Thomas [33]'s meta-analysis, a moderator analysis was performed. This is used to determine whether the relationship between two variables depends on (is moderated by) the value of a third variable [99]. The central idea in this kind of analysis is to identify study attributes ("moderators") which can be applied in order to divide the total set of relevant investigations into two or more subsets, each with at least two elements.

Gerpott and Thomas [33] found through this analysis that, on the one hand, mobile internet usage measurement approach was responsible for the differences in the average correlations. On the other hand, the geographical origin of the samples was also found to have a moderating impact on the MI usage. They compared the samples collected in Europe and the United States of America with those collected in Asia. The results suggested that (1) the used measurement method and (2) the sample origin have an important influence in deciding if a potential factor is significant.

3.4.6 Limitations of the meta-analysis

The results of Gerpott and Thomas [33]'s meta-analysis are definitely valuable inputs for the conceptual model. Compared to traditional vote-counting literature review procedures, the application of a full effect size-based meta-analysis has important advantages according to Gerpott and Thomas [33]. For example, it provides a quantitative synthesis of past findings which accounts for sampling and measurement errors. However, there are still limitations in this kind of analysis. First, it discards valuable results, because quite a number of investigations did not contain the complete information required to enter a study in an effect size meta-analysis. Secondly, it assumes that associations are comparable across studies even though the specific operationalisations for a variable pair are not identical. Operationalisation is a process of defining the measurement of a phenomenon that is not directly measurable, though its existence is inferred by other phenomena. Thirdly, regardless of Gerpott and Thomas [33]'s efforts to comprehensively identify relevant studies, it cannot be excluded that some studies were overlooked or misclassified. Fourthly, due to sample size restrictions, the range of moderators, which Gerpott and Thomas [33] was able to analyse, was narrow and no multivariate assessment of the unique impacts of one moderator, after partialling the influence of others, was possible. To try and compensate for this, other empirical studies, which discuss other potential factors, were also taken into account.

3.5 Cisco forecasting model

The American multinational technology conglomerate Cisco Systems has an ongoing initiative to track and forecast the impact of visual network applications called the Cisco Visual Network Index (VNI). They are able to acquire more accurate forecasts of, among others, mobile (data) traffic due to their cooperation with a large number of service providers and direct data collection. In the latest published Cisco VNI White Paper [32], an example of its methodology for forecasting internet video volumes was explained. This is shown in Fig. 3.4.

Cisco [32] explained that this forecast methodology has been developed based on a combination of its analyst projections, in-house estimates and forecasts, and direct data collection. These projections for broadband connections, video subscribers, mobile connections, and Internet

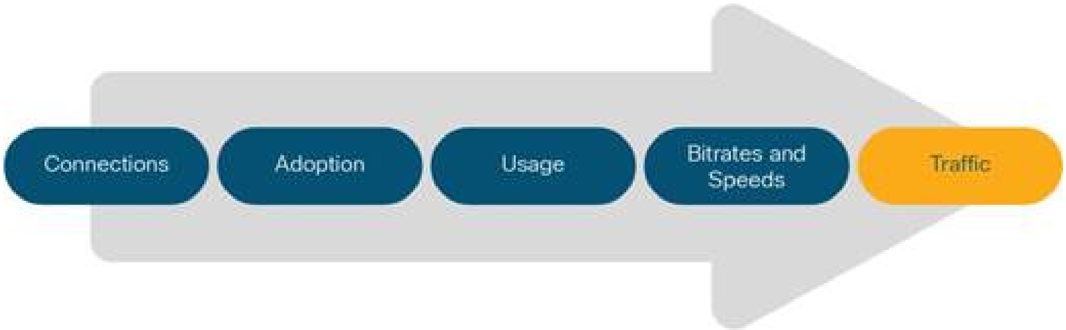

Fig. 3.4: Cisco forecast model for Internet video traffic [32]

application adoption come from a broad range of partners and research institutes. Cisco then layered its own estimates for application adoption, minutes of use, and kilobytes per minute upon this foundation. The adoption, usage, and bit-rate assumptions are tied to fundamental enabling factors such as broadband and computing speed. All usage and traffic results are then validated using data shared with Cisco from service providers.

Compared to which potential factors were found in section 3.4, it can be seen that Cisco also uses fundamental parameters such as adaption and usage pattern. For internet video traffic, they incorporated the network speeds into their forecasting model. Cisco will first estimate the average number of minutes a user will watch a video. This is then multiplied by the available bitrates for the user to get an estimate of the data volumes consumed. This suggests that a higher network speed leads to faster internet and data related services (e.g. video streaming) access for mobile users. Consequently, this would make using mobile data for video streaming more attractive resulting in a higher consumption of mobile data.

Cisco will of course have a more detailed and accurate forecasting model since they have more researchers developing and improving it on a daily basis. For this dissertation, the focus lies on developing a conceptual forecasting model for predicting the roaming data volumes in general.

3.6 Insights obtained from literature review

Through the analysis of existing studies on potential factors influencing the MI usage, important insights were obtained. Instead of only relying on logical reasoning for choosing factors to be included in the conceptual model, some of these are supported through literature. Even with the meta-analysis' limitations as [33] remarked, it is still useful to include the most promising factors (found via this method) in the model, test their significance and analyse how they affect the roaming (data) use. Since few studies focused on the prediction of roaming volumes, additional factors need to be included in the model to account for roaming specifically. These additional factors are inspired by the available datasets and intuitive reasoning. With these insights, a conceptual model will be developed in the next chapter.

Chapter 4

Conceptual model

In this chapter, several approaches for constructing a conceptual model for analysing the roaming data volumes in abstracted economic regions will be described. The input parameters selected for the model are described, these are based on two different inputs. On the one hand, the insights from the literature study obtained in Chapter 3. On the other hand, roaming specific parameters are included based on reasoning and the available datasets. Lastly, the method for predicting the roaming volumes and analysing the impact of each parameter are described.

4.1 Different approaches

The goal of this dissertation is to analyse the roaming volumes, however, there are no further specification on which roaming volumes are of interest. Therefore, it can already be useful to decide which specific part of the roaming volumes will be the focus of the conceptual model. There is a difference in complexity when forecasting different volumes. For example, the forecast of the total yearly roaming volumes will have to deal with the seasonality of each quarter, e.g. in the summer period there is typically more roaming traffic due to more people travelling and going on holiday. Another interesting part is the peak volumes during a specific season or period. In this case, it depends on the country or region because countries such as Spain which has more tourists in the summer period will have different peak (inbound) traffic during that season than e.g. Sweden. On the other hand, the daily/weekly peak volumes will also have some kind of seasonality, e.g. during the day there will be more use than past midnight or differences between a typical weekday and the weekend. Forecasting the average usage per subscriber as given in the BEREC roaming benchmark reports also has its complexity. Here, all the peak volumes are averaged out and can therefore give misleading results which means that forecasting these volumes correctly is more difficult without sufficient data. Depending on which part of roaming volumes the focus lies, additional parameters need to be added in explaining the variability of that part.

Another interesting differentiation is the prediction of inbound versus outbound roaming volumes. As explained earlier in section 2.3.3, the impact of RLAH will be different for operators of sending and receiving countries. The different types of roaming volumes from an operator's perspective were explained in Chapter 2 (see Fig. 2.1). Operators from sending countries are most interested in the forecast of future increases in outbound volumes, because the majority of their roaming traffic is of that type. A large increase means that they have to potentially adjust their retail prices in order to remain profitable and competitive in the roaming market. The same reasoning can be followed for operators from receiving countries. In their case, the inbound roaming volumes are more important and additional capital investments may be required for upgrading their existing network in order to handle the increases in inbound roaming traffic.

Analysing the factors influencing the roaming volumes is important, but this is insufficient to get a view on the real impact of RLAH on operators. The (estimated) roaming volumes serve as an input for the cost model of operators. In such a model, different inputs, other than the roaming volumes predictions, are used to estimate the total cost of providing roaming services for their customers; these include infrastructure costs, wholesale costs, other business operating costs, etc. From the output of this cost model, the actual change in costs (as a result of RLAH) can then be analysed. This is the real impact operators will experience. Since analysing every cost component of the cost model lies outside the scope of this dissertation, this will not be discussed further on.

In the remainder of this chapter, a model focusing on estimating the roaming data usage per user will be developed. The reason for this choice is to maintain the generality of the model. Also, predicting the usage per user is more accurate compared than predicting e.g. the total volume of all users. Another reason is partially due to the format in which the data was available, because the roaming volumes were given in average monthly use per subscriber (more specifically the outbound volumes). The data will be analysed further in detail in the next chapter (section 5.1.2). The factors that will be defined serve as an initial step in developing the model. These are based on literature insights and intuitive reasoning. This model can be extended or modified for analysing specific parts of the roaming volumes such as the peak volumes.

4.2 Submodel: roaming usage per subscriber

The model discussed in this section focuses on the roaming data usage per subscriber, more specifically the outbound volumes. From the available datasets, it is very difficult to analyse the inbound roaming volumes, because data for all parameters of the model needs to be collected from all visiting users in order to estimate the inbound volumes. Therefore, it is chosen to predict the outbound volumes from a specific country's point of view. The conceptual model is visualised in Fig. 4.1.

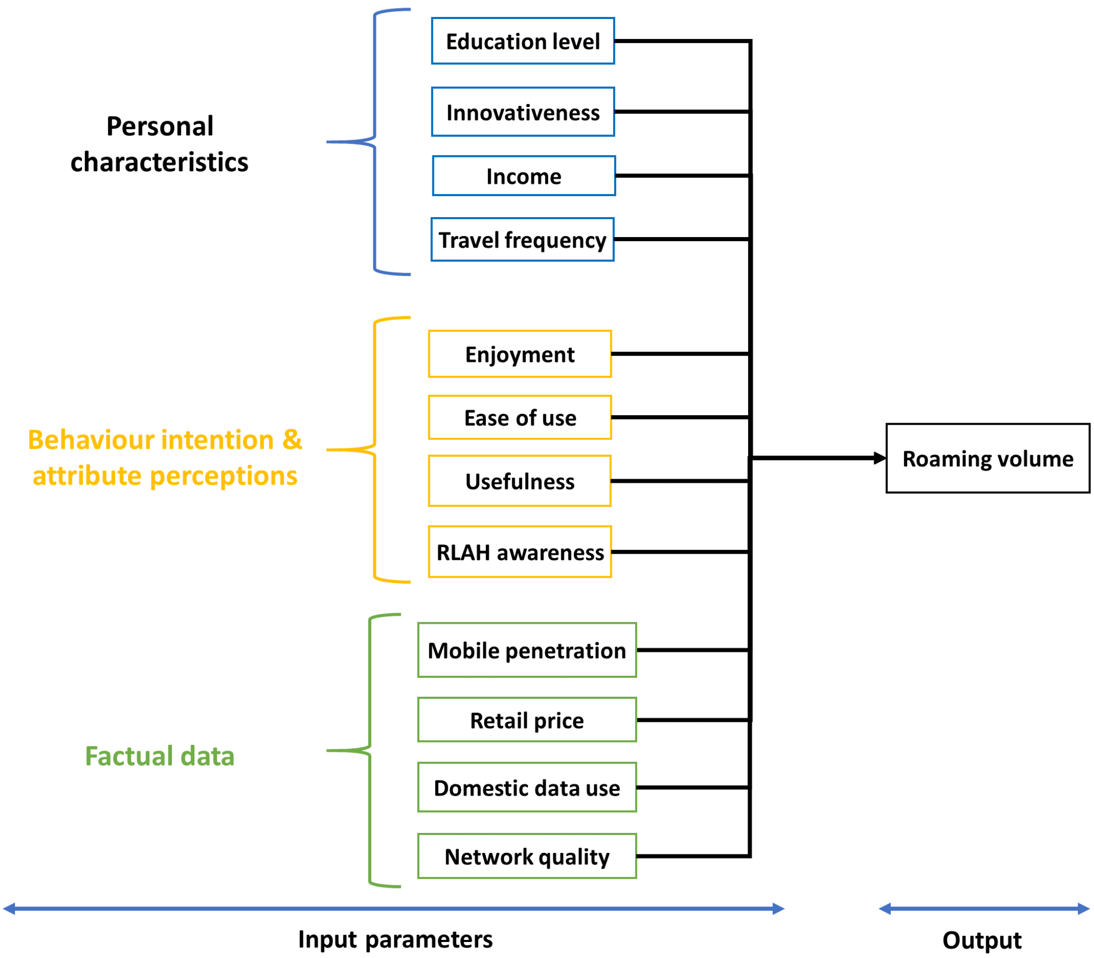

Fig. 4.1: Conceptual model containing potential factors influencing the roaming use

4.2.1 Output variable

The output variable of this model is a user's roaming data use per time period. The reason for using this unit of measurement is based on the available datasets. If the roaming use is expressed in another unit of measurement, then that can be used as output as well. Regardless of the data, the output of the model is always the roaming use. Whether it is the total use of an entire country/region or the per subscriber use.

To estimate this output variable some historic data must be used to fit or train the model with the selected input parameters of this model. The most accurate way to collect these data is to cooperate with the mobile operators, because they have access to objectively measure data from their network. In the ideal scenario, data from all parameters should be collected from the same population of subscribers in order to have more reliable data. It can be expected that operators will be reluctant to share this information publicly due to, among other things, their customers' privacy concerns.

4.2.2 Input variables

The input variables/parameters for predicting the roaming volumes are described below. These are retrieved from the insights of the literature study and the available datasets. The parameters are grouped into three categories discussed below.

Personal characteristics parameters

In this category, four parameters are included: (1) Education level, (2) Innovativeness, (3) Income, and (4) Travel frequency. The first three parameters are selected from the insights gained in section 3.4.2 while the fourth one is added for the purpose of roaming specifically.

The parameter *Education level* showed to be influencing the mobile internet usage [74], and by extension, the roaming use as well. One could argue that there would not be a large difference in the education level of EU citizens because the EU countries are considered to be developed economies. Therefore, the difference between countries as a whole should not be significant. However, since the output of the submodel is focused on the roaming usage per user, the difference in education level between user groups can have a larger difference, e.g. users who completed higher education versus those who did not. Therefore, it should be useful to include this parameter in the model. To get more accurate estimates, the model should be applied on different user groups to determine whether this parameter is actually useful in explaining the roaming use.

Innovativeness denotes the degree in which an individual will try out new technologies. This parameter was found to be useful in past studies and thus chosen to be included in this model. However, the question raises again whether the difference in EU countries are significant. In the absence of data for this parameter, it is difficult to determine whether the innovativeness should be left out or not. However, different user groups, e.g. younger versus older people, could show a larger difference in innovativeness. For example, elderly people may be reluctant to try out new technologies such as smartphones or mobile internet. The usefulness of this parameter should again be verified by applying the model on different user groups.

The parameter *Income* also showed some influence on the MI use [33]. This refers to money received by a person or household over some period of time, it includes wages, salaries, and cash assistance from the government. It is reasonable to assume that this parameter is useful because a person with a low income will most likely not choose a very expensive mobile subscription; this means that the price per unit of data will be higher and the included data volume per month is lower. On the other hand, a person with a higher income will not necessary use more mobile data, because this is dependent on his or her personal habit of use. If that person has a more expensive subscription plan, then a higher usage is more likely. Following this reasoning, the income can influence the mobile data use, and by extension the roaming use. It is expected that there is a significant difference across European countries. Since there was no data available for the average income per person, another related parameter was used to illustrate the differences between countries, i.e. the wealth per adult. This refers to the stock of assets held by a person or household at a single point in time. These assets may include financial holdings and saving, but also real estates; therefore, wealth can generate a part of the income for a person. The wealth per adult (given in US dollars) in different countries are summarised in Appendix A.4. It can be seen that there is a clear difference between the wealthy to very wealthy countries which lie in the Western and Northern parts of Europe (e.g. Luxembourg = 412.13$ per person) compared to the poor to very poor countries which primarily lie in Eastern Europe (e.g. Lithuania = 24.60$ per person). As a result, this difference is expected to be reflected in the roaming use as well.

The last parameter in this category is the *Travel frequency*. It must be mentioned that this denotes the travel frequency to countries outside the home country of the user and to regions where RLAH is active. This is an intuitive chosen parameter to include, because the more a person travels, the higher the probability a person will use roaming services. It could also be useful to differentiate between users who only travel for personal enjoyment and those who travels a lot more for work purposes. The latter will typically have other usage patterns compared to the former user group.

Behaviour intention and attribute perceptions parameters

In this category, there are four parameters included in the model: (1) Enjoyment, (2) Ease of use, (3) Usefulness, and (4) RLAH awareness. The first three are inspired by the analysis in section 3.4.3. The last one is intuitively chosen based on the available datasets.

The first parameter is the *Enjoyment* of users when they are using mobile internet. It must be mentioned that this denotes the enjoyment of using mobile internet in general and not necessary roaming related. However, the more enjoyment a user gets from using this, the more likely a user will keep using it while roaming. This is a rather vague parameter because the enjoyment that one person experiences, is different from another. For example, some users only read news articles on their devices while others primarily watch online videos for enjoyment. Therefore, the first group will consume a lot less data compared to the second group. As a consequence, both groups could attain the same level of enjoyment, but the consumed data volume is very different. This suggests that dividing the user's populations into different groups is useful in order to get more accurate weights of each parameter. So why was this not done in previous studies analysed in Chapter 3? The reason for this is because of the limited number of users participating in those studies. The participants were mostly from the same (academic) environment of the researchers which meant there was a risk that the results were not representative for the entire population of end users. Besides the actual data volume that is required for a user's enjoyment, the network quality also affects the experience of using the service. For example, when the network coverage is less in certain regions then the user will get less satisfaction in using mobile data as a result of the longer loading times to access webpages or streaming videos. The group which consumes less data is affected less by this problem compared to the other group which consumes more data. Either way, a bad network quality will definitely have a negative influence on the enjoyment of a user. As a result, the (average) network quality has a direct influence on the enjoyment of the end user. Therefore, both enjoyment and network quality should be included to model the effects of both parameters. When reliable data is available and further analysis shows that this parameter is not of interest, then this can be left out of the model, but not in this initial phase.

The second parameter is (perceived) *Ease of use*, i.e. how easy users find it is to use mobile data on their phone. Since the mobile penetration in EU countries is quite high, it is reasonable to assume that most phone users know how to enable mobile internet on their phones, however, this is not the only important aspect. As with the previous discussed parameters, different user groups may have other opinions on this parameter. For example, younger generations who grew up using the internet on electronic devices (smartphone, tablet, etc.) will find it much easier to use compared to elderly people who may not know how to use web browsers on their phone to access the internet. Therefore, it is useful to include this parameter. On country level in the EU there may not be much difference for this parameter, but when the model is applied to different user groups there could be a larger difference. Therefore, is it still useful to include this factor into the model and perform further analysis based on reliable data. It is then easy to understand the importance of collecting data from a representative population, i.e. users from all ages, genders, user profiles, and backgrounds.

The third parameter is (perceived) *Usefulness*, i.e. how useful users think roaming (data) is for them. An example of usefulness is when a user can use mobile internet to handle work-related issues such as answering e-mails or attend a meeting via video calls. The more functions mobile data can provide for end users, the more likely they will be inclined to use it while travelling. The difference will again be noticeable between different user groups, e.g. users who travel frequently for work and those who do not.

The final parameter in this category is the *Roam Like at Home awareness*. This parameter is inspired by the large-scale survey done by the European Commission shortly after the inception of roam like at home in 2017 [100] with the aim of collecting relevant data for the analysis of the impact of RLAH. From the data that was collected (see Appendix A.9), it can be seen that for all EU countries, more than half of the respondents in each country is aware of RLAH. The differences between is not as extreme as the Wealth per adult parameter discussed earlier. However, there is still a noticeable difference between the countries. For example, France (59%), the United Kingdom (63%), and Romania (63%) showed that they have the least number of respondents who are aware of RLAH compared to countries such as Belgium (85%) or the Czech Republic (84%). Since it is reasonable to assume that a higher awareness would lead to more roaming (data) use in general, it is useful to include this parameter in the model.

Factual use data

In the last category, there are also four parameters included: (1) Mobile penetration rate, (2) Retail price, (3) Domestic data traffic, and (4) Network quality. These are primarily inspired by the available datasets.

The first parameter is the *Mobile penetration rate*, this is defined as the number of SIM cards in a country. Note that this does not refer to the actual number of mobile phone devices since SIM cards can also be used in, e.g. tablets to use mobile data. It is usually presented as a percentage and can exceed 100% if the number of SIM cards in a country is larger than the actual population number. The higher the mobile penetration, the higher the probability of people having access to mobile internet. Therefore, the inclusion of this parameter should be useful.

The second parameter is the *Retail price* for using mobile data, more specifically the price per unit of data. The reason for choosing this parameter is derived from two factors found in Gerpott and Thomas [33]. There it was found that the tariff plan of the end user and its monetary value were of influence on its mobile internet use. The retail price incorporates both of these factors. In case of the tariff plan, the more gigabytes of monthly data a subscription plan offers, the lower the price per gigabyte of data. So, a user who uses lots of data will automatically choose this kind of subscription plan because this is the most advantageous in terms of costs for the user. However, the retail price should also be linked to the income of the end user and its living expenses when comparing users from different countries. For example, it can be seen from the available retail prices for mobile subscriptions in general, which was given as a relative value between one and four compared to the European average (see Appendix A.8), that Belgium has a value of three which means relatively expensive while Lithuania has a value of one which means inexpensive compared to the European average. When only focusing on the retail price, one would get the impression that Lithuanian mobile users do not have to pay much for their mobile use. However, this is not entirely true, because when the income parameter is also taken into account (discussed above and represented by wealth per adult in US dollar), it can be seen that Belgians have on average 313.05$ per person while Lithuanians only have 24.60$. In that sense, the retail price must be placed in the perspective of the country itself meaning that there is an interaction between income and the retail price when both are used in one model. From a Belgian user's point of view, the Lithuanian mobile subscriptions are far cheaper, but this does not necessary mean that the same is experienced by Lithuanian end users. For them, their mobile subscriptions may be quite expensive due to their lower income/wealth. One could question whether the retail price alone also has an influence on the data use. Intuitively, a lower retail price per unit of data would encourage the user will probably use more data and vice versa. The positive influence of a lower retail price on a user's (domestic) data use is also observed from the available datasets. Hence, the retail price is useful to include as one of the parameters, but the income factor is also important due to the interaction effect of these two parameters which is why it is also included.

The third parameter is the domestic data *Traffic*. The domestic use serves as an indication for the roaming use, because it is expected that an end user's habit will carry over when travelling. In past studies, this data was mostly collected via self-reporting of users which had larger error margins. With more accurate data from mobile operators, the actual effect of this parameter can be better estimated.

The last parameter in this category is the *Network quality*. As explained earlier, since an end user's enjoyment is directly influenced by the network quality, it should be useful to include the average network quality in a certain country/region in the model as well.

4.3 Analysis method

Once sufficient data is collected, the goal is to (1) test which parameters are significant and (2) what contribution each parameter has on the output.

Dependent on available data and the level of differentiation possible, there are a wide variety of methods for predicting the roaming volumes and analysing the influencing factors. There are the conventional regression techniques to fit the available data. With very large datasets operators can also apply big data methods to better analyse the datasets and use, e.g. machine learning algorithms to predict the roaming use of their subscribers. From the available datasets (see section 5.1), it can be seen that only a limited number of samples were available for further analysis. Even for a standard linear regression analysis, there is a large risk of over-fitting due to the number of explanatory variables that is inserted into the model to predict the roaming volumes. Regardless of the chosen analysis method, the model will have low statistical power due to the small sample size. With this in mind, the multiple linear regression method was selected for the estimating the roaming data use in this dissertation. The reason for this choice is because a regression analysis is used to describe the relationships between a set of independent variables (= inputs) and the dependent variable (= output). From output of the regression analysis it can then be determined which variables are statistically significant meaning that they should be kept in the model. This analysis produces a regression equation where the coefficients represent the relationship between each independent variable and the dependent variable [101] which is the desired output of the model. The goal is to identify the relationship between the potential influencing factors (= independent variables) and the roaming volume (= dependent variable). The regression equation can then be used to make predictions about the dependent variable. Of course, the assumption in a linear regression is that each dependent variable, i.e. the input parameters, has a linear relationship with the output variable. This is not necessarily valid for each parameter. There is also the possibility to perform a non-linear regression instead, but in that case one must also include the expected non-linear function (quadratic, cubic, etc.) as input as well. Since the estimation of this polynomial is already a complex procedure together with the small sample size and available time frame, this non-linear fitting method was not explored further. As mentioned before, the focus is more on the conceptual part instead of purely data focused, therefore, only one fitting method was used to analyse the available data. The multiple linear regression will be performed using the statistical computing software R.

The overall fit of the model should be analysed based on the obtained R^2 and adjusted R^2-values [102]. The R^2-value, which lies between 0 and 1, indicates how much variability the independent (input) variables of the model can explain compared the total variability of the studied phenomenon, i.e. the roaming volumes. The closer this value is to 1, the better the model fits the data. Besides the normal R^2-value, the adjusted R^2-value should also be analysed since this value penalises the addition of too many input variables, which contribute little to explaining the total variability, to the model [103]. Since the R^2-value alone cannot determine if the coefficient estimates and predictions are biased, the residual plots must also be analysed [104].

Besides the overall model fit, the individual predictors (i.e. the input parameter/variables) should also be tested for their significance [102]. The output of the regression analysis contains t-values for each of the predictors and its corresponding p-values. The lower this p-value, the more significant a predictor variable is, meaning that it should be kept in the model. In the literature (e.g. Montgomery [103]), a p-value of 0.1 is taken as threshold for deciding whether a predictor variable is statistically significant. If the p-value is below 0.1 for a predictor, then that predictor is assumed to be statistically significant. Other values for p are the 0.05 or 0.01 significance level [103].

In linear regression, there are a number of assumptions which must be fulfilled. First, all variables are required to be multivariate normal which essentially means that all variables should be normally distributed such that any linear combination of the variables is also normally distributed. This is useful in extending the central limit theorem to multiple variables. The central limit theorem (CLT) is a theorem about independent random variables, which roughly says that the probability distribution of the average of independent random variables will converge to a normal distribution, as the number of observations increases [103]. This is a very useful theorem because it essentially tells us that you do not need to know the real distribution of an independent random variable, which does not necessarily have to be normal, but you can always approximate it by the normal distribution when the sample size is sufficiently large. Secondly, there should be a linear relationship between the input variable(s) and output variable. Thirdly, the residuals of the regression model should be normally distributed. Fourthly, there should be little to no multicollinearity in the data. Multicollinearity occurs when independent variables are too highly correlated with each other which is undesired in a regression model. The last assumption is homoscedasticity, this means that the variance around the regression line is the same for all values of the predictor variable(s) [105].

Chapter 5

Results and analysis

With the input parameters defined for the model, the data collection can now be done. This was followed by applying the regression techniques to the collected data. In this chapter, the conceptual model will be applied to the EU-28 countries to analyse the mobile roaming data use of end users. The reason for only focusing on these countries is because RLAH is implemented in these regions. The methodology is as follows: the collected data is first screened for potential outlier countries; these are then removed for the further analysis. Next, a multiple linear regression with the collected data is performed in the statistical software R followed by a discussion of the results and insights gained related to the impact of RLAH.

5.1 Data collection and analysis

For this master's dissertation, only public databases were available, the majority of them were from EC-related organisations. The main sources for objectively collected data were the quarterly International Roaming BEREC Benchmark Data Reports [10], [19]–[22]. These have aggregated a wide range of data from the NRAs in EEA countries. Other useful sources were the statistical office of the European Union *Eurostat* [106] and the Digital Scoreboard database of the EC [107]. Shortly after the introduction of RLAH, the EC organised a large-scale survey to gather new information about the EEA citizens' view on RLAH [100]. This provided some useful data for the input parameters as well. Lastly, the Mobile broadband prices report of the EC [108] was also consulted.

5.1.1 Collected data

Since only public databases were consulted, not all twelve parameters defined in the conceptual model (Fig. 4.1) could be collected. A summary of the available parameters is given in Table 5.1.

Table 5.1: Collected parameters of the conceptual model

Variable	Available?	Replaced by	Source
Education level	No	-	-
Innovativeness	No	-	-
Income	No	Wealth per adult	Global Wealth Databook [109]
Travel frequency	Yes	-	Eurobarometer survey [100]
Enjoyment	No	-	-
Ease of use	No	-	-
Usefulness	Yes	-	Eurobarometer survey [100]
RLAH awareness	Yes	-	Eurobarometer survey [100]
Mobile penetration	Yes	-	Digital Scoreboard database [107]
Retail price	Yes	-	Mobile broadband prices report [108]
Domestic data use	Yes	-	Roaming benchmark reports [21]
Network quality	No	-	-
Roaming data use	Yes	-	Roaming benchmark reports [21]

In the *personal characteristics* category, only the Travel frequency parameter could be retrieved from the Flash Eurobarometer survey [100]. The Income parameter was not directly obtainable and therefore replaced by the Wealth per adult parameter for which data was available. The wealth per adult and its relationship with the income were explained earlier in section 4.2.2, therefore, it is useful to represent the income with this parameter. This data was collected from the 2017 Credit Suisse Research Institute's Global Wealth Databook [109]. The other two parameters (education level and innovativeness) could not be obtained from the available databases. Hence, these two factors cannot be included in the regression analysis. Consequently, the potential impact of these factors could not be estimated.

In the *behaviour intention and attribute perceptions* category, data for two parameters was available: Usefulness (1) and RLAH awareness (2). The data was retrieved from the Eurobarometer survey [100]. The Usefulness parameter was collected via the question of: 'Do you think RLAH is beneficial for you or your family members and friends?'. The RLAH awareness was collected via the question of: 'Do you know that RLAH is in effect?'.

In the *factual use data* category, data for all three parameters were available. The BEREC benchmark reports provided quarterly data for domestic and roaming volumes for each of the EEA countries. The unit of measurement for these volumes is the average amount of use (in megabytes or gigabytes) per subscriber per month. It must be mentioned that the roaming volumes defined in this context denote the roaming use in the EEA, the roaming use in the rest of the world is not included.

Besides data volumes, the mobile penetration rate was also available through the Digital Scoreboard database of the European Commission [107]. This penetration rate was given as the number of SIM cards per one hundred people.

Finally, data for the retail prices was collected via the Mobile Broadband Prices in Europe report [108]. This report mentioned limitations of the presented results. For example, the collected data was only a snapshot the retail prices at the time of collecting. Another limitation was that only the two or three largest MNOs in a country were included. This means that although the included MNOs had a market share of at least 70%, it was still possible that a smaller MNO or MVNO had a less expensive offer for one or more usage baskets. Due to the large number of different tariff plans available in each country, the OECD defined different usage baskets and relate a certain price to these. However, it was noted that these baskets may not be a good reflection of the actual usage. The report used a number between one and four to represent a country's retail price (not for mobile data specifically, but for mobile services in general) by comparing it to the EU average. A relative value of one means a low retail price, a two means relatively lower than the EU average, a three means relatively higher than the EU average, and a four means a high retail price. The same earlier remark can be made, i.e. the differentiation in a country's retail prices is lost by using only one value to represent the price.

The roaming and domestic data volumes from Q4-2017 are used in the regression analysis further on, these were retrieved from the BEREC benchmark reports. The other parameters were collected from various sources, there was no differentiation between quarters for these, therefore, the yearly data was used instead. The collected parameters and their original unit of measurement are summarised in Table 5.2. The exact values of the data points used in the following regression analysis can be found in Appendices A.1 to A.9.

5.1.2 Data analysis

In this subsection, some preliminary analysis on the collected data is done to filter out potential outliers before the regression analysis. The method used for determining outliers is based on calculating the quartiles and constructing a boxplot. This method was first introduced and described by American mathematician John Tukey [110] back in 1977. Since then, it has been widely used to determine potential outliers in datasets. The first (Q1) and third (Q3) quartiles are calculated and the difference between these two values is called the interquartile range (IQR). The lower acceptable bound (LB) is calculated as $Q1 - 1.5 \cdot IQR$. The upper acceptable bound (UB) is calculated as $Q3 + 1.5 \cdot IQR$. A data point is assumed to be a potential outlier when it lies below the lower bound or above the upper bound. The use of

Table 5.2: Variables used for the regression analysis

Variable	Unit of measurement
Roaming data use	monthly average per subscriber in megabytes [21]
Domestic data use	monthly average per subscriber in gigabytes [21]
Travelling in EU-28	% respondents who travelled in the last 12 months to EU-28 countries [100]
Wealth per adult	US dollar [109]
Usefulness	% respondents who thinks RLAH is beneficial for them [100]
Turn off mobile data	% respondents who turn off mobile data on their devices when travelling in the EU-28 countries [100]
Mobile penetration	active SIM cards for voice or data per 100 people [107]
Retail price	relatively represented with a number between 1-4 [108]
RLAH awareness	% respondents who is aware of RLAH [100]

1.5 times the IQR is a rule of thumb developed by Tukey, his reasoning was that this is valid when the data has a normal distribution. The different parts of a boxplot are visualised in Fig. 5.1.

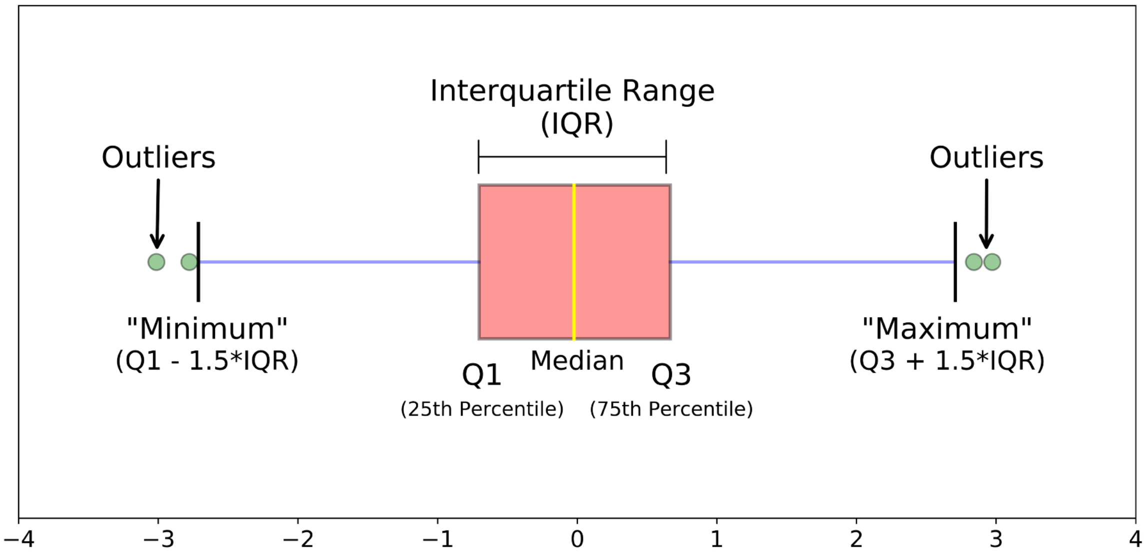

Fig. 5.1: Different parts of a boxplot [111]

Roaming data use

The data is summarised in Appendix A.1, these are the outbound roaming data volumes. It can be seen that there is a large variation in roaming data use between the different countries in the EU. The lowest value is 83.57 MB/month for Slovakia while Cyprus has the highest value of 688.11 MB/month. The calculated lower bound is negative which means every value lies above the LB. The upper bound is 434.98 MB/month. The roaming use in Cyprus and Sweden are higher than this upper bound, therefore, these two countries are considered to be potential outliers. As a result, the decision was made to remove these two countries from the regression analysis.

Domestic data use

The data is summarised in Appendix A.2. In this case there is also a large variation on country level. Greece only uses 0.66 GB/month while Finland uses 16.84 GB/month. The domestic data use is also related to, among others, the retail price. A lower retail price per unit will have a positive influence on the consumption of data. The calculated lower bound is negative which means no data point lies below this bound. The calculated upper bound is 8.06 GB/month. It can be seen that domestic data use is Austria and Finland are above this bound, therefore, the decision was made to exclude these countries from the regression analysis as well.

Travelling in EU-28

The data is summarised in Appendix A.3. The lowest travel frequency was observed for Greece (12%) while Luxembourg had 85%. This is still a large difference which means depending on the impact of this parameter, this could be useful in explaining the variation in roaming data use. None of the values lie outside the lower and upper bounds which means no potential outliers were observed in this dataset.

Wealth per adult

The data is summarised in Appendix A.4. The lowest wealth per adult was observed in Romania (20.30$) while the highest was in Luxembourg (412.1$). The central and eastern EU countries have a lower to very low wealth in general. The EU countries in the North and West seem to be wealthier based on this dataset. None of the values can be considered to be potential outliers for this parameter.

Usefulness

The data is summarised in Appendix A.5. The variation in usefulness is smaller than the previous than the parameters discussed above, however, it is still a noticeable difference on country level. At least 60% of the respondents in every EU country found RLAH to be beneficial for them. The lowest measured value was in France (59%) while the highest was in Ireland (91%). There are also no values which could be considered to be outliers for this parameter. All values are within the bounds; therefore, no potential outliers are present in this dataset.

Turn off mobile data when travelling

The data is summarised in Appendix A.6. The percentage of people who turned off mobile data while travelling ranges from 11% (Malta) to 50% (United Kingdom). It can be seen that there is still a considerable difference between countries. All values are within the bounds; therefore, no potential outliers are present in this dataset.

Mobile penetration

The data is summarised in Appendix A.7. In terms of mobile penetration there is a large difference in the EU countries. The minimum was measured in Hungary with only 48.7 subscriptions per 100 people compared to Poland with 154.9 subscriptions per 100 people. There is also no clear distinction in different groups of countries, e.g. more wealthy (West Europe) versus less wealthy (Eastern Europe) countries. It is reasonable to assume that when there are more subscriptions per a certain amount of people, the probability of using mobile services (including mobile data) will increase for that country. Therefore, this factor is expected to have a positive influence on the mobile use and by extension the roaming (data) use. Although there are large differences between countries, none of these values lie outside the acceptable bounds, therefore, no outliers are assumed.

Retail price

The data is summarised in Appendix A.8. The retail price is only given relatively as a number between one and four and uses the EU average as a reference point. More detailed data was not available. The retail price should not be viewed on its own. Instead, it should be linked to the living costs and income of a country's citizens. For example, from a wealthy country's perspective a certain country with a low retail price (compared to the EU average) seems beneficial, however, this is not necessarily true for the mobile users of that country. They could have low wealth in general which means that the "low" retail price in their perspective is more expensive. Therefore, it is essential to include a parameter to represent their income as well. In this case, the wealth per adult is used.

RLAH awareness

The data is summarised in Appendix A.9. The awareness in the EU countries is at least 50% which is positive sign. The lowest values were measured in Greece (53%) while the highest ones were observed in Luxembourg (87%). It is reasonable to assume that this will only increase in the future and become more homogeneous in all countries. In the meantime, the current differences on country level should have some impact on the roaming use as a result of RLAH. All values are within the acceptable bounds; therefore, no potential outliers are present in this dataset.

5.1.3 Data limitations

Since the (outbound) roaming volumes were given as the average use per subscriber per country, this means that there are only 24 data points of this parameter available (the four outlier countries: Cyprus, Sweden, Austria and Finland were removed) for the regression analysis. With only one data point per country, there is no way to differentiate the data volumes between different user types/groups in a country. As a consequence, it is difficult to estimate the average roaming use accurately for end users of a specific country with only one value per country. Using only one data point to represent a country also means that only one value can be used to represent the other input parameters in the regression model.

There are also potential problems when combining datasets gathered from different populations. The data presented in the BEREC benchmark reports were gathered from a much larger population than the data from the EC's Eurobarometer survey. The four parameters which were retrieved from the EC's survey [100] are: (1) Travelling, (2) Usefulness, (3) Turn off mobile data, and (4) RLAH awareness. The EC explained in this survey that the methodology used to select potential participants ensured a relatively representative population for each country. Essentially, they used call centres in each country where they randomly generated phone numbers to call, this process was repeated until they reached about one thousand participants in each country. They said that the results should be representative for each country with a few percentages of error margin. Although this seems sufficient to gather an initial view for each country, the data collected in this way may not be sufficient when used as predictors for the roaming volumes. There are a few reasons for this. First, the roaming volumes were obtained from a different and much larger population than the participants of the survey. BEREC accumulated these data from mobile operators (MNOs and MVNOs) in each country. Together with the domestic data use, these are the most reliable datasets. In the ideal situation, the other parameters would also be collected from the same population, however, this is very difficult to achieve in reality and not true in this case. So, the mixture of datasets from two different populations already leads to inaccuracies. The survey data is most likely not representative for the (much larger) population from which the roaming volumes were collected. Another potential problem with the survey data is that the participants consist of people who are actually interested in the roaming subject in general. Since the completion of all questions in this survey would take around five minutes based on the number of questions they would have to answer, it is then reasonable to assume that lots of people (without interest in roaming) would refuse to waste their time taking part in this survey. Consequently, that group of mobile users were already missing or under-represented in the survey data which would lead to misleading results. Other potential problems with survey data were discussed earlier in section 3.2.1 such as the risk of over-reporting data etc.

Since a theoretical approach was used to develop the model, it is almost inevitable that some input parameters would not be available when only public databases are consulted. Therefore, other available data was analysed and added to the regression model as input variables if these could potentially have an influence on the roaming use. One parameter was added using this approach, i.e. the amount of people who turned off their mobile data services when travelling. The question raises why some potential factors of influence (on the roaming use) were not found in the literature. The reason for this is because the exponential increase in roaming use (within the EEA) only happened after the introduction of RLAH in 2017. As a result, few studies focused on forecasting the roaming volumes with potential influencing factors as inputs. These past studies focused on mobile internet use in general and not specifically on roaming. Therefore, it is reasonable to assume that roaming specific factors could be overlooked or simply neglected. Even if a study focused on forecasting mobile data volumes in general (e.g. Hong, Dong, Chen, *et al.* [112]), the focus was not so much on identifying which factors influenced the data use, instead the model fitting was more important. They used different fitting techniques such as time-series forecasting and heuristics (genetic algorithms, simulated annealing). However, these only tried to determine coefficients for a model which fits one dataset, i.e. the total mobile data use, without further analysing what factors influenced these volumes. As a result, the literature inputs, i.e. the factors which were found to be useful in past studies, for this model must first be verified using reliable data to determine whether these are currently still important for the roaming data volumes. The search for an appropriate method to forecast these volumes is still a current topic of research.

5.2 Regression analysis

From the data analysis in the previous section, it was decided to remove some countries from the original EU-28; these are: Cyprus, Sweden, Austria and Finland. This resulted in a total of 24 data points for the use in this regression analysis. The average (monthly) roaming data use of a subscriber is fitted in function of the input parameters. The roaming data use per subscriber (= output) is seen as the dependent variable while the input parameters are assumed to be independent variables (= predictors). The goal of this analysis is to determine the correlations between the predictors and the output variable in order to analyse what impact each parameter has on the output.

In this analysis it was decided to represent the parameters relatively compared to its EU-28 average, because this allows an easier interpretation of the results. Therefore, the values of each variable are converted first before using them in the regression model. For example, a value of 1.5 after conversion means that the variable is 50% above EU-28 average while a value of 0.7 means 30% below the average. There is a certain risk of interpreting the data in this way. For example, the retail price in a certain country could be very low compared to the EU average, however, the mobile users in that country may still find the price quite high due to their lower income and living expenses.

Each predictor is then inserted independently to fit the output. This allows a pairwise interaction analysis of the output variable with individual predictors. This is equivalent to a simple linear regression between each of the predictors and the output variable. This step is useful since the number of samples is limited. According to VanVoorhis and Morgan [113], in order to test the overall fit of the model with k predictor variables, a minimum sample size of $50 + 8k$ is recommended. If the significance of k individual predictors needs to be tested, then VanVoorhis and Morgan [113] recommends a minimum sample size of $104 + k$. The conceptual model contains a total of twelve input parameters (Fig. 4.1). Therefore, the minimum required sample sizes for testing the model fit and the individual predictors should be 146 and 116 respectively. With the EU-28 countries as target population, there should be at least six samples per country. These are the minimum required sample sizes since no interaction effects are currently included as predictors. Additional samples are needed when interaction effects need to be included as well.

It must be mentioned that the predictors used in this regression model are not necessarily independent. For example, the retail price is most likely affecting the used mobile data. However, due the sample restrictions, the predictors were assumed to be independent and additional interaction effects were neglected in this regression analysis. The risk, as a result of this assumption, is that these potential interaction effects cannot be modelled/explained by this simplified model.

5.2.1 Simple linear regressions: pairwise comparisons

Each predictor variable in Table 5.2 is now individually tested as an independent variable to predict the roaming volumes as the dependent variable resulting in eight simple linear regressions. From the regression output, the predictor's regression coefficient, p-value, and its standardised coefficient were analysed.

The regression coefficient denotes the change in the dependent variable (= output) when the predictor is increased/decreased with one unit. If this coefficient is positive, then there is a positive correlation between the predictor and the output while a negative coefficient means that there is a negative correlation. For example, it is reasonable to assume that the roaming use is negatively correlated with the retail price, because a higher retail price per unit of data will most likely lead to a decrease in roaming use and thus a lower amount of data units consumed. The p-value of a predictor is used to determine whether a predictor is statistically significance, if this is the case then the predictor should be included in the model. The standardised regression coefficient of a predictor, which will be called the beta-value from now on, was also analysed. This denotes the change in standard deviation of the dependent variable when the predictor's standard deviation is increased/decreased with one unit. Therefore, it can be used to analyse a predictor's impact on the output variable. The sign of this coefficient shows the correlation between the dependent variable and its predictors similar to the regression coefficients. The reason for analysing the beta-values is the following: when a predictor that is strongly correlated, but has a low beta-value, its impact on the output variable is limited. As a result, it does not contribute much in explaining the variability of the output variable meaning that keeping such predictors in the model is not very useful; removing these and adding other more impactful variables may be more interesting. On the other hand, a predictor that shows little correlation, but has a high beta-value will have a significantly larger impact on the output variable. In this case, it is worthwhile to further investigate why this predictor has a larger impact and perhaps the inclusion of interaction effects containing this variable is necessary to improve the regression model.

The results of the simple linear regressions are summarised in Table 5.3. The different Intercept values of the different simple linear regressions are not included in this table.

Table 5.3: Results of the simple linear regressions

Predictor	Coefficient	p-value	Beta-value
Domestic data use	0.2180	0.1430	0.3078
Travelling in EU-28	0.2322	0.1833	0.2811
Wealth per adult	0.1685	0.1019	0.3420
Usefulness	0.0336	0.9641	0.0097
Turn off mobile data	0.1156	0.6054	0.1110
Mobile penetration	0.8156	0.0062	0.5424
Retail price	−0.4038	0.0111	−0.5091
RLAH awareness	0.0730	0.9128	0.0236

The signs of most coefficients in these pairwise comparisons make sense, i.e. the correlation relationships between the predictor and the output variable seem reasonable. For example, a higher retail price leads to a lower roaming use. The only exception is for the turning off mobile data when travelling parameter. In this case, the model suggests a positive correlation between the roaming use and this parameter as predictor. This is highly unlikely since one cannot expect the roaming use to increase when more users are turning off data services on their devices. Hence, removing this predictor in the multiple regression model seems useful in obtaining better and more comprehensive results. However, this does not necessarily mean that this predictor is not useful in reality, because this dataset might just be a wrong match with the dataset of the roaming volumes, i.e. the data for this parameter was taken from a non-representative population.

Based on the p-values, it can be seen that there are three significant predictors: (1) Mobile penetration, (2) Retail price, and (3) Wealth per adult. Here, the normal criterium for significance (p-value = 0.1) is used. Although the Wealth per adult has a p-value slightly over 0.1, this parameter is also assumed to be significant. The other predictors cannot be called statistically significant according to these results. However, due to the limited sample size, the statistical power of the model is limited. As a result, there is a possibility that a predictor, which is actually important in reality, is overlooked due to the model not being able to show this in its p-value. Even for one predictor the minimum required sample size is equal to 58 according to [113]. Therefore, the conclusions drawn from these results must be interpreted with caution. Common sense must also be used to reason whether the results make sense. The two other predictors which are the closest to being statistically significant are Domestic data use (p-value = 0.1430) and Travelling (p-value = 0.1833). The least significant predictors according to these results are Turn off mobile data, RLAH awareness and Usefulness.

From comparing the beta-values, it can be seen that the most significant predictors are: Mobile penetration and Retail price. These also show the largest impact on the output variable with beta-values of 0.5424 and -0.5091 respectively. Wealth per adult (beta-value = 0.3420), Domestic data use (beta-value = 0.3078) and Travelling (beta-value = 0.2811) have similar impact. The predictors with the lowest beta values are: Turn off mobile data, RLAH awareness, and Usefulness. The order of these predictors in terms of impact is similar to the previous paragraph. The signs of the beta-values correspond to the signs of the regression coefficients.

Based on these regression outputs, it seems that the Turning off mobile data predictor is not fit to be used for this particular dataset of roaming volumes due its positive regression coefficient and beta-value. One potential explanation is that this predictor is indeed not a good one to include in the model. Another reason could be due to the datasets itself, i.e. the data was collected through surveys. The problems regarding these survey data, discussed earlier (section 5.1.3) could be a possible explanation of the very low impact of RLAH awareness on the roaming use. Intuitively, the awareness is expected to have a larger influence on one roaming use, however, this could not be seen in the regression outputs. The low impact of the Usefulness parameter is more acceptable, since this parameter was retrieved from past studies and not necessarily relevant for this current dataset. Also, the variation for this parameter was also smaller compared to the other parameters as discussed in section 5.1.2.

The observations from this analysis should only be used as an estimate of the potential effects/impacts of a predictor, because errors are still possible as a result of the data limitations.

These should be verified once more data points become available.

5.2.2 Multiple linear regression

From the simple linear regression results, it was observed that only the following three predictors were significant according to the p-values:

- Mobile penetration
- Retail price
- Wealth per adult.

However, the following parameters with p-values above the significance level of $p = 0.1$ were chosen to be included in the multiple regression model as well:

- Domestic data use
- Travelling in EU-28
- RLAH awareness.

The reason for this choice was because some parameters, which are important in reality, could be missed out by the model, i.e. not achieving statistical significance based on the p-values, due to the data limitations. For example, the Domestic data use was not seen as statistically significant in the previous subsection due to its p-value, however, it might be important in reality. Therefore, one must not solely focus on the p-values, but use some intuitive reasoning as well for selecting parameters. Hence, some parameters, were kept in the multiple regression model despite them not showing statistical significance in the pairwise comparisons earlier on. As a result, six out of the eight analysed predictors in section 5.2.1 were selected to be included in the multiple regression model. As previously stated, the sample size is already insufficient for a reliable simple regression analysis, therefore, a more complicated multiple linear regression model using the same number of samples will also not be a good fit. Nevertheless, an attempted was made to fit the output variable with six predictors and see what insights could be gained from the model results. The predictors' regression coefficients, p-values, and beta-values are analysed similar to the previous subsection. Additionally, the overall model fit will be evaluated based on the R^2-values together with the general model assumptions described in section 4.3. The selected predictors and the corresponding results of the multiple linear regression analysis are summarised in Table 5.4.

The signs of most predictors' regression coefficients remain the same compared to the pairwise comparisons (section 5.2.1), except for Domestic data use and RLAH awareness. The signs of those two are different in this case. The observed negative correlation between the Domestic data use and the Roaming use seems counter-intuitive. The same can be said about the

Table 5.4: Results of the multiple linear regression

Predictor	Coefficient	p-value	Beta-value
Intercept	1.0998	0.146	-
Mobile penetration	0.7346	0.103	0.4885
Retail price	−0.1589	0.421	−0.2003
Wealth per adult	0.0363	0.789	0.0737
Domestic data use	−0.0265	0.885	−0.0373
Travelling in EU-28	0.2769	0.381	0.3352
RLAH awareness	−1.0772	0.283	−0.3487

change in correlation between the RLAH awareness and the roaming use. The reason for this could lie in the fact that some other potential predictors are missing, e.g. interaction effects between current predictors. One of these effects could be the interaction between the mobile penetration and the domestic data use, e.g. the more SIM cards per person, the more likely a person will use mobile services and by extension mobile data. In the pairwise comparisons, each of the predictors was used individually to fit the output (= roaming use). However, when these predictors were combined into a single model, the regression signs or coefficients of these individual predictors can change as observed here. By analysing the absolute values of these coefficients, it can be seen that RLAH awareness is responsible for the largest change in roaming use compared to the other predictors. Mobile penetration has the second largest coefficient value. The remaining predictors have significantly lower coefficients. It is remarkable that the RLAH awareness now has a much larger coefficient in this model while its contribution was almost non-existent in the simple linear regression. This led to the belief that the RLAH awareness does have a larger influence on the roaming use when used in a multiparameter model, which was intuitively expected, however, this could not be derived from the values of the simple linear regression. Although its negative correlation is highly unlikely, it does suggest that its effect should not be neglected.

The p-values show that only Mobile penetration is statistically significant in this regression model. It is remarkable that even the p-value of the Intercept does not achieve statistical significance. This suggests that the model is not a good fit. Most notable are the higher p-values for domestic data use and wealth per adult. As mentioned before, these two predictors are not necessarily independent of each other, therefore, these high p-values could suggest that these two parameters should not be used individually in the model, but rather as an interaction term. The same could be said about the possible interaction effect of the retail price, wealth per adult, and domestic data use. Constructing these interaction terms is rather difficult since these predictors are not factors. In a factorial analysis, the interaction terms can be obtained by multiplication of two factors and inserting them in the model. However, with continuous predictors, one must find a good way to construct these terms. Should the values of these predictors be multiplied with each other or should they be added together instead? Trial-and-error is involved in attempting to find the suitable interaction term. Some of these combinations (e.g. retail price multiplied with wealth per adult) were attempted, unfortunately, these did not lead to better results. This was expected due to the sample size restrictions and potential inaccuracies in the survey data as discussed earlier. Therefore, it was chosen to not further investigate the construction of these interaction terms because there was insufficient data to verify if these were good ones to include or not within the time frame of this dissertation. The main takeaway from these results is that interaction effects between parameters are most likely to be present based on the (large) changes in regression coefficients and p-values from the simple linear regression to the multiple linear regression. For example, RLAH awareness which had little influence when used on its own became much more important when used in a multiparameter model. According to these results, only Mobile penetration is significant. However, more parameters are expected to affect the roaming use as found in literature (see section 3.4). One predictor will most likely not be able to explain sufficient variation of a complex phenomenon such as the roaming use.

Next, the beta-values were also analysed. Mobile penetration remains the predictor with the largest impact on roaming use. The same change in importance of RLAH awareness was also observed through its beta-value together with its unexpected change of sign, which is counter-intuitive, compared to the simple linear regression output. The impact of domestic data use and wealth per adult also drastically decreased in this multiparameter model which again suggests that interaction terms are most likely to be present.

The model has a R^2-value of 0.4667 (adjusted $R^2 = 0.2785$) and its p-value is 0.0658. In theory, a good model should at least have a R^2 value of 0.95, preferably 0.99 or higher. However, a low R^2-value is not inherently bad, because this model tries to predict human behaviour instead of a physical process. In some fields, it is entirely expected that the R^2-values will be low. For example, any field that attempts to predict human behaviour, such as psychology, typically has R^2-values lower than 0.5 [104]. However, important insights can still be gained from the predictors showing significance. For example, in this case Mobile penetration seems to be have consistent importance in predicting the roaming use.

It is difficult to interpret the predictions (compared to the actual roaming use) of this particular model because these can be misleading. The reason for this is due of the following problems. In the simple linear regressions, it was found that the RLAH awareness was highly insignificant (regression coefficient $= 0.073$, p-value $= 0.9128$). This factor was still used in the multiple regression model because it was assumed to be of greater influence on the roaming use. This was then also suggested from the outputs of this model. Its coefficient suddenly became much larger ($= -1.0772$) and its correlation with the roaming use changed as well. It also became closer to being statistically significant (assuming $p < 0.1$ to be significant) since the p-value dropped to 0.283. The change from a positive into a negative correlation of this parameter means, according to this particular model, the less users are aware of RLAH, the more roaming data they would use. This is highly unlikely and counter-intuitive. Consequently, there is no way of rationally explaining these predictions regardless whether these are under- or overestimations of the actual roaming use. However, one should not focus too much on the deviations of the predictions compared to the actual values. Instead, the focus should lie on the changes in the regression coefficients, p-values, and beta-values of the parameters from the simple to the multiple regression model. These give an indication whether a predictor has a potential influence on the roaming use or the potential presence of interaction effects. The changes in output for RLAH awareness suggest that there are interactions with one or more of the other predictors. Due to data limitations, it is difficult to determine which interaction effects are present/important. One possible interaction could be between the RLAH awareness and travel frequency. Other predictors in this model also showed noticeable changes. The domestic data use changed from a positive regression coefficient into a negative one ($0.218 \rightarrow -0.0265$) which seems counter-intuitive since the domestic use should serve as an indication, i.e. a positive influence, for the roaming use assuming that the habit of the user carries over when travelling. Its potential significance also became less, the p-value changed from 0.1430 to 0.885. Similar changes were also observed for the wealth per adult parameter. Although the positive correlation of this parameter with the output variable did not change (which seemed reasonable), its coefficient did decrease a significant amount ($0.1685 \rightarrow 0.0363$). It also became a statistically non-significant parameter based on its p-value which was initially 0.1019, but changed to 0.789. All these changes suggest, as discussed earlier, that interaction effects are present and perhaps some of these should be used as predictors instead of only the individual predictors that is used in the current model. For example, the interaction between wealth per adult and retail price (as explained previously), or the interaction between those two parameters and the domestic data use. The only consistent parameter was mobile penetration since its coefficient remained similar and it was also still significant based on its p-value.

The multiple regression model discussed above contains counter-intuitive negative correlations between the output (i.e. the roaming use) and two of its predictors (i.e. the RLAH awareness and domestic data use). The RLAH awareness had a large influence on the predictions due to its large regression coefficient (-1.077). For the reasons explained above, it was decided to remove the RLAH awareness and the domestic data use of this model with the aim of getting better results, i.e. predictions for which the contribution of each predictor could be explained more rationally. The roaming volumes are now fitted without these two predictors and the results of this updated model are summarised in Table 5.5.

The regression coefficients (and their signs) of all predictors make sense. Mobile penetration

Table 5.5: Results of the updated multiple linear regression

Predictor	Coefficient	p-value	Beta-value
Intercept	0.4883	0.289	-
Mobile penetration	0.4962	0.136	0.3300
Retail price	-0.2397	0.171	-0.3023
Wealth per adult	0.1275	0.243	0.2589
Travelling in EU-28	0.0098	0.957	0.0118

has the largest coefficient value and will have a larger impact on the predictions since it absolute value (0.4962) is at least two times larger than the other predictors. It is remarkable that the influence of the travelling predictor is much smaller compared to the simple linear regression results, there it was found that it had a larger regression coefficient and beta-value. However, its effect might be lowered when used together with other predictors like in this model. None of these predictors seem to be statistically significant according to the p-values, but mobile penetration is still the closest to being statistically significant with its p-value of 0.136. The signs of the beta-values correspond to those of the regression coefficients which is expected. The absolute beta-value of mobile penetration is decreased while that of the retail price is increased compared to the previous model making their impact on the roaming use more similar.

This updated model has a R^2-value of 0.4247 which is slightly lower than the previous model, however, its adjusted R^2 is 0.3036 which is slightly higher meaning that it this particular model is a better fit for the used datasets. Looking at the adjusted R^2 is actually better since this value also penalises the addition of too many predictors which have little contribution in explaining the variation of the output variable. The p-value of the new model is 0.0264 which is also lower than before which suggests that the updated model is a better fit.

Lastly, the assumptions of the updated multiple linear regression model, described in section 4.3, are also investigated. The first one is that all variables are normally distributed. This is an assumption that was made in the beginning. The second one is that there should be a linear relationship between the predictors and the output variable. When there are only two variables involved, this can be examined using a scatter plot. However, since there are four predictors included in the model, scatter plots are not possible anymore. Instead, the Residuals vs. the fitted plot (Fig. 5.2) can be used to investigate this assumption. If there is a strict linear relationship between each predictor and the output variable, the red line on this plot should be perfectly horizontal. This is clearly not the case for this model. Therefore, a non-linear relationship between some predictors and is output is most likely the case. Another assumption, which can also be checked using this same plot, is the homoscedasticity of the residuals. If this assumption is fulfilled, the data points should be equally spread out around the dotted horizontal line (Residuals = 0.0), this means that no patters should be present. In this case, the plot seems to be acceptable.

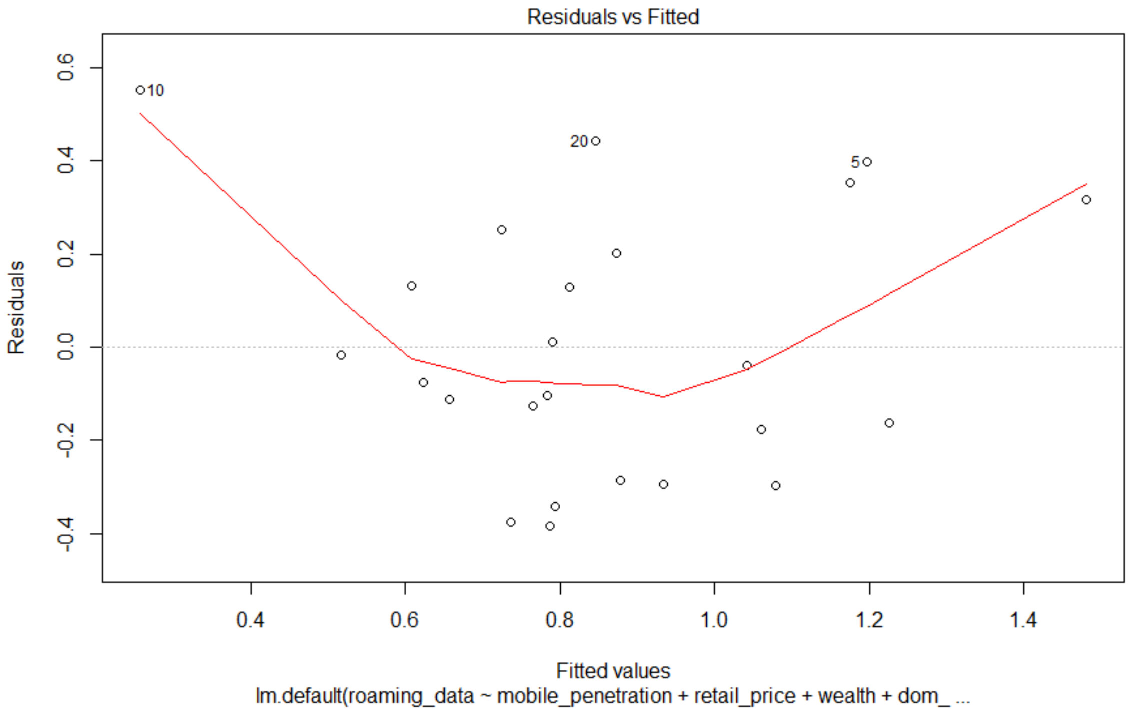

Fig. 5.2: Residuals vs. fitted values plot of the multiple linear regression model

The next assumption is that the residuals of the regression model should be normally distributed. This can be examined using the Quantile-Quantile plot (Fig. 5.3). This is a probability plot, which is a graphical method for comparing two probability distributions by plotting their quantiles against each other. In this case the normal distribution is on the x-axis while the residuals are on the y-axis. If the residuals are perfectly normal, then the data points should lie perfectly on the diagonal line. Some residuals are further removed from the diagonal, but this seems to be acceptable. Another way of verifying this assumption is to use the Shapiro–Wilk test of normality [114]. The null hypothesis of this test is that the data is normally distributed. The obtained p-value of this test is 0.2933 which is larger than the 0.1 significance level. Hence, the null hypothesis cannot be rejected in this case meaning that the residuals are considered to be normally distributed.

The last assumption is the possible multicollinearity in the data. This can be examined by inspecting the Variance Inflation Factor (VIF) [103]. When different variables are uncorrelated with each other, their VIF should be equal to 1. According to Montgomery [103], a VIF lower than 5 indicates a small multicollinearity and is still assumed to be acceptable. Any value

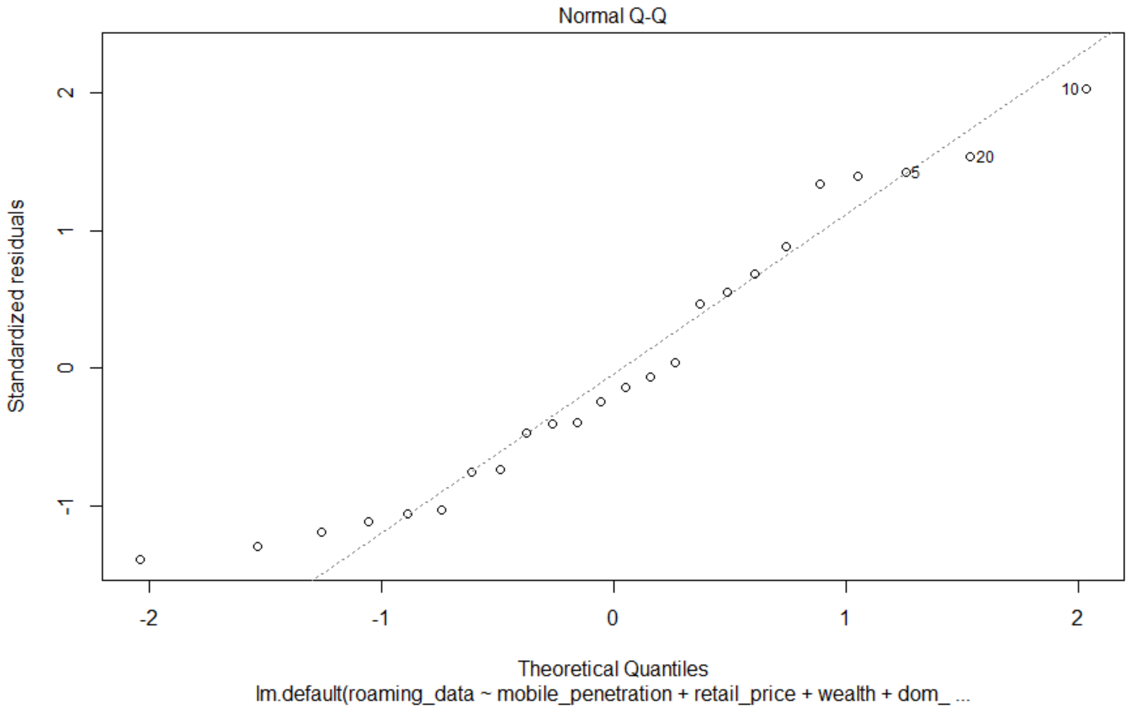

Fig. 5.3: Quantile-Quantile plot of the residuals

above 10 is considered to be unacceptable. The VIF values of the predictors are summarised in Table 5.6. It can be seen that these values are acceptable, however, depending on the context of the model, sometimes a VIF of two can already be a problem. In this case it is assumed to be acceptable.

As mentioned before, one must not put too much emphasis on these values, the main takeaways should be the most promising predictors according to these results and the presence of potential interaction effects. According to the multiple regression results, the three predictors which look the most promising are: (1) mobile penetration, (2) retail price, and (3) wealth per adult. These are intuitively easy to understand. A higher mobile penetration means there

Table 5.6: Variance Inflation Factor of the predictors

Predictor	VIF
Mobile penetration	1.483
Retail price	1.489
Wealth per adult	1.524
Travelling in EU-28	1.557

are more SIM cards per person, therefore, the higher the probability that people will use mobile services and by extension mobile (roaming) data. It is also expected that the retail price and the wealth per adult (and their interaction effect) will most likely also affect the data use. Hence, these three parameters will most likely be important in reality and should be included in future models. Additionally, the two predictors which were removed in the updated multiple linear regression model (i.e. RLAH awareness and domestic data use) should also be further investigated to determine whether these should have been kept in the model. These were removed due to their unexpected change in impact when used together with the other parameters. This can only be done when more data becomes available.

5.2.3 Comparison with model predictions

The predictions of the updated multiple regression model are plotted against the actual roaming volumes to see how much these deviate from the actual volumes. The EU countries: Cyprus, Sweden, Austria, and Finland are excluded since these were assumed to be potential outliers as found in the initial data analysis. One must also note that the predictions here are produced by only four predictors (mobile penetration, retail price, wealth per adult, and travelling in EU-28) which means that the potential impact of other parameters are neglected. The comparisons can be found in Fig. 5.4.

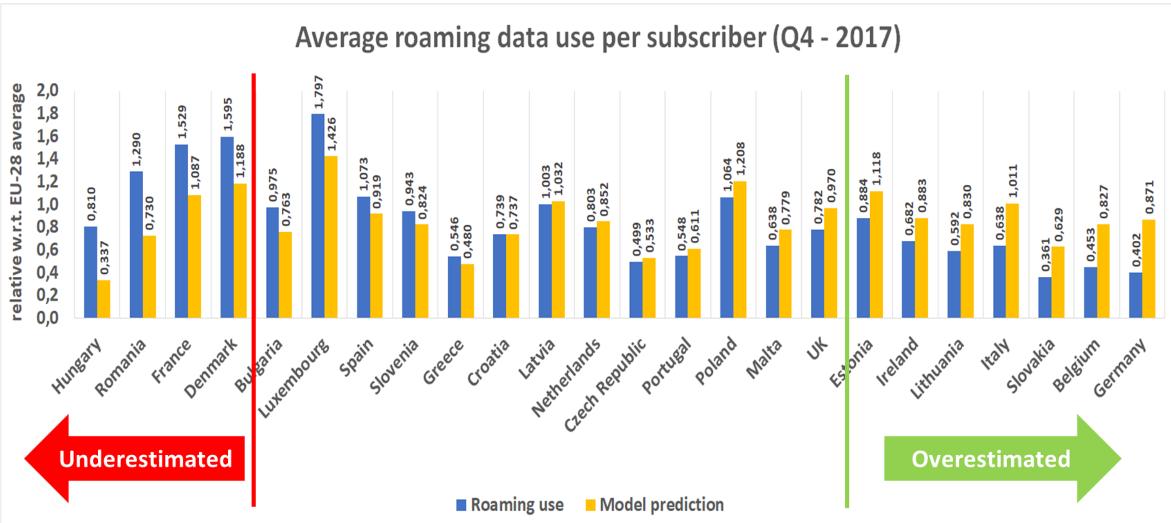

Fig. 5.4: Comparison of regression model predictions with the actual roaming volumes

The countries on the left side of this figure are the ones where the roaming use is underestimated by the multiple regression model, while the countries on the right side are overestimated. It was decided to label a prediction acceptable when its deviation from the actual value stayed below 25%. Using this criterium, the countries with acceptable predictions are summarised in Table 5.7.

Table 5.7: Countries where the roaming volumes were acceptable

Country	% difference with actual value
Bulgaria	−22
Luxembourg	−21
Spain	−14
Slovenia	−13
Greece	−12
Croatia	0
Latvia	+3
Netherlands	+6
Czech Republic	+7
Portugal	+11
Poland	+14
Malta	+22
United Kingdom	+24

The four underestimated countries are summarised in Table 5.8. The most underestimated country is Hungary where the actual roaming use is 58% more than its prediction. Romania was underestimated by 43% followed by France (29%) and Denmark (26%).

The overestimated countries are summarised in Table 5.9. Most notably are the severe overestimations of Germany, Belgium, and Slovakia. The regression model predicts more than twice the actual roaming use in Germany. The deviations in the overestimated countries are considerably higher compared to the underestimated ones. Where the maximum percentage deviation was 58% previously, this has become more than 117%.

The mobile penetration and retail price have a large impact on these predictions since their regression coefficients are the largest. Also, the potential impact of the excluded parameters is not taken into account in these predictions. The model under- or overestimates eleven out of the twenty-four countries considered in the regression analysis. The underestimated countries

Table 5.8: Countries where the roaming volumes were underestimated

Country	% difference with actual value
Hungary	−58
Romania	−43
France	−29
Denmark	−26

Table 5.9: Countries where the roaming volumes were overestimated

Country	% difference with actual value
Estonia	+26
Ireland	+30
Lithuania	+40
Italy	+58
Slovakia	+74
Belgium	+83
Germany	+117

do not share specific characteristics, the same holds for the overestimated ones. For example, in the underestimated countries, there are the less wealthy countries (Hungary, Romania) but also the very wealthy countries (France and Germany). Also, their retail prices and mobile penetration, which have the most impact on these predictions, also vary a lot. The same observations can be made in the overestimated countries. Consequently, it was not possible to group countries together with similar characteristics or define abstracted regions based on the results of this model. The large impact of mobile penetration and retail price together with the absence of the excluded parameters could be an explanation why there are some many countries overestimated.

It can thus be concluded that this regression model is unable to explain sufficient variation in the roaming use. This is intuitively expected since the literature and the regression results suggest that more parameters (which include interaction effects) are needed to predict the roaming volumes more accurately. The results also reinforce the belief that interaction effects of parameters could also influence the roaming use, e.g. when an individual is wealthier, then that person could be less affected by the retail price and still chooses to roam more regardless of the higher price. Hence, more complex forecasting methods may be necessary to model these interaction effects.

5.3 Summary of the regression analysis

The conceptual model developed in Chapter 4 was the starting point of this analysis. Not all input parameters of the model could be collected from the available databases. For the parameters that were collected, only 28 samples were available for each parameter (one value for each of the EU-28 countries). This has limited the validity of the obtained results in the regression analysis, because the sample size was insufficient to test the predictors' statistical significance. Therefore, the conclusions drawn in this chapter must be interpreted with caution and common sense.

The pairwise comparisons (simple linear regressions) showed that the correlation relationships between the predictor and the output variable seemed logical. The three parameters which showed significance were: (1) Mobile penetration, (2) Retail price, and (3) Wealth per adult. It was remarkable that the RLAH awareness had a very low influence on the roaming use (based on its regression coefficient of 0.073) and was highly non-significant (p-value = 0.9128). This could be a consequence of the data limitations.

Although only three parameters (mobile penetration, retail price, and wealth per adult) were statistically significance in the pairwise comparisons, three other parameters (RLAH awareness, travelling in EU-28, domestic data use), which were assumed to be influential, were added to the multiple regression model as well, resulting in a total of six predictors. The output of this model showed some counter-intuitive results. The RLAH awareness suddenly had the largest regression coefficient, however, the most surprising was the change in correlation (i.e. from positive to negative) with the output variable. The same change in correlation was observed for domestic data use as well. In an attempt to counter this problem, these two predictors were removed from the multiple regression model. The updated model results showed more intuitive results, i.e. expected correlations between the predictors and the output variable. Here, none of the predictors were statistically significant according to the p-values. The three predictors which looked the most promising were: (1) mobile penetration, (2) retail price, and (3) wealth per adult. Their importance is intuitively easy to understand. These will most likely be important in reality, hence, these should be included in future models. Other takeaways from these results are: (1) the presence of potential interaction effects of predictors when used together in a multiparameter model and (2) statistical significance was achieve for three parameters in the pairwise comparisons. These observations should be verified and further investigated when more data becomes available. Also, these are conclusions drawn from fitting the outbound volumes, it would be interesting to verify these with the fitting of the inbound volumes.

Chapter 6

Conclusions and future work

The goal of this master's dissertation was to perform a multiparameter analysis on the roaming data volumes in the EU countries and by extension abstracted regions in context of the RLAH legislation. The research questions that this dissertation aims to answer were: which parameters have an influence on the consumed roaming volumes (1) and what is the expected impact of RLAH on a country/region (2).

The methodology used in this dissertation is as follows. First, necessary background information, related to the general mobile roaming concepts (1) and the evolution Europe's roaming legislation (2), was gathered through a literature review. Next, potential influencing factors were retrieved from existing studies related to an end user's intention to use mobile internet services. With these insights, the most promising factors were then selected to be used in the conceptual model. Next, the available databases were consulted to retrieve useful data followed by an initial analysis to identify potential outliers. Then, a multiple linear regression was used to fit the roaming volumes with the available data. From the obtained results, the impact of RLAH in different countries were estimated.

The focus of the conceptual model was to identify potential influencing factors on the roaming data use. The reason for this was because data traffic contributes the most on a mobile operator's network load. Therefore, additional factors related specifically to data use (e.g. travelling habits or turning off mobile data when travelling) were intuitively chosen to be included in the model. Next, the choice of analysing the outbound or inbound roaming volumes was made. Theoretically, this model can be applied to analyse both in- and outbound volumes. The more difficult part is how the model parameters can be gathered from all incoming mobile users in the case of inbound volumes. Therefore, it was decided to focus on analysing the outbound roaming volumes instead.

A total of twelve potential factors were defined as input parameters in the conceptual model. The next step was to collect data for these parameters. For this master's dissertation, only public databases were available, this resulted in limitations regarding the availability of certain model parameters. After the data collection, simple linear regressions were performed first, followed by selecting the most promising parameters and combining them in a multiple linear regression model. In section 6.1, conclusions are drawn based on the results of this model and the research performed in the dissertation in general. Lastly, some future work topics are proposed in section 6.2 from which researchers can continue to investigate in order to further analyse the roaming and RLAH topic in general.

6.1 Conclusions

The developed conceptual model in this dissertation only focused on analysing outbound roaming data volumes. This is partially due to the limited data that was available. The other inputs, from which the potential factors were retrieved, were previous publications (academic, research institutes, BEREC, etc.) related to mobile internet use. It was observed that few studies focused on analysing potential factors of influence on the mobile (roaming) data use. This is partially understandable since roaming data use in the EEA only started to increase exponentially from 2017 onwards (since RLAH) based on the yearly total data use published in the most recent BEREC international roaming benchmark data reports. So, before the introduction of RLAH, there was less focus on a more detailed analysis of these roaming volumes. In past studies, which focused on forecasting the (domestic) mobile data use in general, the focus was not so much on identifying which factors influenced the data use, instead the model fitting was more important. These studies used different fitting techniques such as time-series forecasting and heuristics (genetic algorithms, simulated annealing). However, these only tried to determine coefficients for a model which fits one dataset, i.e. the total mobile data use, without further analysing which factors contributed to these volumes. As a result of the limited studies focused on (domestic) mobile data use specifically, the literature inputs, i.e. the factors which were found to be important in past studies, for this model must be verified using reliable data to determine whether these are still important for the roaming data volumes.

There was data collected for eight input parameters to be used in the regression analysis. Based on the initial data analysis, four countries (Cyprus, Sweden, Austria and Finland) from the original EU-28, which were potential outliers, were removed from the regression analysis. From the results of the simple linear regressions, it can be seen that that three parameters were significant based on their p-values. These are: Mobile penetration (p-value = 0.0062), Retail price (p-value = 0.0111), and Wealth per adult (p-value = 0.1019). The other five parameters were not found to be statistically significant. However, due to the data limitations leading to potential inaccuracies for certain parameters and limited statistical power of the model, some of these parameters which were not found to be statistically significant, could potentially be important in reality. Therefore, three of these parameters, which were intuitively chosen, were also selected to be used in the multiple regression model together with the statistically significant parameters. These are: Domestic data use, Travelling, and RLAH awareness. When the parameters were combined in a multiple linear regression model, the results showed some unexpected results, i.e. the change from a positive to a negative correlation for the parameters RLAH awareness and domestic data use. This suggested that there were potential interaction effects present between the parameters. Due to data limitations and the available time frame, it was difficult to perform a quantitative analysis on these interaction effects. Based on intuitive reasoning, it can be expected that there is a relation between the retail price and the income of the mobile users and therefore a potential interaction effect of importance. Another potential interaction is one between the travelling frequency and the RLAH awareness since the more an end user travels, the more he or she would come in contact with information about roaming. Although the Domestic data use did not achieve statistical significance in the regression models using the current data, it is reasonable to assume that it would be useful to include this parameter in future analyses, because the domestic usage habit of a user should give an indication on its roaming use when the awareness of RLAH further increases and the majority of the users realises roaming (in the EEA) will not bring any extra costs (when the fair use limit is respected of course). In an attempt to obtain predictions for which the contribution of each parameter could be explained more rationally, these two predictors were removed in the updated model. The new model results suggested that none of the predictors were statistically significant. The non-significance of the individual parameters suggested the presence of potential interaction effects between input parameters. Nonetheless, the multiple regression results suggested that the three most promising parameters are: Mobile penetration, Retail price, and Wealth per adult. Their importance is intuitively easy to understand as explained in section 5.2.2. This corresponds to the results found in the simple linear regressions earlier on. However, whether the other parameters were actually non-significant remain uncertain. This must be verified when more data becomes available.

When the linear relationship assumption between the predictors and the roaming use of the regression model was examined, it was observed that a non-linear relationship for some of the predictors was more likely the case. Hence, a non-linear fitting approach might be a better option. However, this was not further investigated due to three reasons. First, a non-linear regression approach requires the input of the estimated relationships between the fitted variable and its predictors, e.g. linear, quadratic, cubic, etc. This additional estimation is already a complex process since the actual important parameters have not been known yet at this point. Together with the fact that the limited number of samples would not give reliable results, it seemed that this would further complicate an already difficult problem. Secondly, the focus of this research was to identify potential influencing factors using a more conceptual approach instead of purely data focused. Thirdly, other forecasting methods, e.g. time-series or machine learning, focus on one dataset only, i.e. the roaming volumes, and try to fit this dataset the best way possible instead of identifying the potential relationships between the roaming volumes and the chosen predictors (which is what is actually desired in this research). For these reasons, it seemed more useful to use the simpler multiple linear regression method because the predictors could be inserted and tested without further inputs that a non-linear regression model required. In this way, an initial view on the potential important parameters could be obtained.

Next, the predictions of the multiple regression were then compared to the actual roaming volumes. A difference of less than twenty-five percent was assumed to be acceptable. There was no clear differentiation observed in the countries that were either over- or underestimated, e.g. it could not be said that less wealthy countries tend to be underestimated by this model. The same holds for the countries with acceptable predictions. As a result, grouping countries together or defining abstracted regions containing countries with similar characteristics were not possible. Eleven out of the twenty-four countries considered in this model were either over- or underestimated. Consequently, this regression model proved to be insufficient in explaining the variability of the roaming volumes, which was expected due to the potential non-linear relationship between some of the predictors and the roaming use as observed from verifying the model assumptions. It was also difficult to group countries into certain groups or regions based on the actual roaming volumes, because countries with similar roaming use did not show similar characteristics. For example, it could not be said that less wealthy countries (e.g. Slovakia with a monthly roaming use of 84 MB/subscriber) tend to use less roaming because Belgium is a wealthy country but the monthly roaming use is also only 105 MB/subscriber which is considered low compared to other EU countries. Another grouping of countries could be obtained by dividing them into inbound and outbound countries, however, since only data for outbound volumes was available, it could not be determined whether this is a good way of diving the countries. Analysing the inbound volumes might give a better way of grouping countries together, e.g. touristic versus non-touristic regions. The four countries (Poland, Finland, Estonia, and Lithuania) which BEREC mentioned as exceptions should be monitored more closely in the future in order to improve the sustainability of those operators since as derogations were granted to major operators due to very low retail prices and high wholesale traffic asymmetry compared with other member states.

With the aim of analysing the impact of RLAH for different operators and regions, the question raises whether the forecast should focus on the average roaming use instead of peak use in different time periods or seasons. The average value might be misleading since it conceals the very heavy users from the normal users which is currently the case with only one data point per country. This is a potential problem because according to the literature, a small part of the users (i.e. the heavy ones) are responsible for the majority of the data use. Consequently, special attention should be paid for these users from an operator's perspective. It could be more meaningful for operators to focus on the evolution in roaming use for different user groups (i.e. light, normal, heavy) in order to obtain a more accurate view on which part of users will most likely be responsible for the largest increase in roaming use and by extension the change in costs. Also, different types of operators will experience a different impact (see section 2.3.3), e.g. operators from net receiving countries such as Spain are more interested in the evolution of the inbound volumes since this is the majority of their roaming traffic. For these reasons, it is highly unlikely that one model will be sufficient to model the complex phenomenon of roaming use. Therefore, further differentiation of user groups and traffic types are necessary to forecast the roaming volumes more accurately. The models for each user group might contain different input parameters because each group has different characteristics. The same goes for the inbound versus the outbound models. The possible improvements are discussed in the next section.

It can thus be concluded that the mobile roaming use, especially data services, is a complex phenomenon which cannot be analysed thoroughly without sufficient and reliable datasets. Hence, more complex forecasting methods, e.g. non-linear approaches, and further differentiation between user groups and traffic types might be necessary to better model the roaming use. Nonetheless, this research contributed to the roaming topic by developing a conceptual model with the parameters supported by literature. The results of the limited regression model suggested the potential importance of three parameters. It also reinforced the belief that using one model will most likely not be sufficient. The insights gained in this research and the future work suggestions in the next section serve as a good starting point from which future researchers can perform a better analysis on roaming and obtain better estimates without having to start from scratch. As a result, a better view on the evolution of roaming use in different regions and countries can be obtained from which the real impact of RLAH in terms of cost changes for mobile operators can be seen.

6.2 Future Work

In this section, some future topics on the roaming and RLAH will be described. These are necessary steps in order analyse the real impact of RLAH on different types of operators. These are topics which could not be investigated in this dissertation due to, among others, the data limitations and the available time frame.

First, more data samples must be collected from a representative population, i.e. users with different user profiles, of all ages, etc. This is necessary in order to obtain more reliable results. To achieve this, an advanced collaboration between mobile operators is needed to provide enough data samples for each parameter of the conceptual model. The European Commission could use its detailed databases which collects quarterly data from various NRAs. However, it is unlikely that these will be sufficient, because there are parameters which cannot be objectively measured. Therefore, subjective parameters, such as travel frequency or RLAH awareness, need to be measured through surveys. Mobile operators should also select a representative population from its subscribers and organise a large-scale survey to gather personal characteristics and behaviour intention data which cannot be objectively measured. All parameters should be collected from each individual of the population. This information should be centralised by, e.g. providing it to the EC. Once the EC has these sets of complete data from each of the European countries, the multiple regression method suggested in this dissertation can for example be used to obtain more reliable results and gain better insights into the parameters' importance/impact.

Secondly, a number of studies in the literature ([78], [84], [89], [92], [94], [115]–[119]) observed that data use across mobile customers are highly skewed, i.e. a very small portion of "heavy" users causes a large fraction of the mobile data carried over the mobile operator's network. Therefore, it would be useful to split the users into three types/categories: 'light', 'medium', and 'heavy' users. In this way, a separate model for each type of users will result in better estimates for each group, because each parameter will have a different coefficient/impact for each type of users. This was, e.g. something that could not be investigated in this dissertation because the roaming volumes were only available as a monthly average per subscriber. The average values will even out the peak volumes consumed by the 'heavy' users. For this reason, the regression model was not a good fit for these average volumes. The current model is focused on the outbound volumes and the parameters included are not necessarily important for all users. Some could potentially become totally irrelevant for a certain type of user. Therefore, it could be useful to differentiate between different user groups in order to obtain more accurate estimations instead of trying to estimate the usage of all users together in one model. Other ways of dividing end users are also possible, e.g. users who pays for their own subscription versus those who have a mobile subscription provided by their work which they do not have to pay themselves. For the latter, the retail price parameter will most likely not have an influence at all on their mobile (roaming) use.

Next, different models can be explored as well. As mentioned above, the current model is focused on outbound volumes, the logical next step would be to apply the model to analyse the inbound volumes as well. When both volumes are available, a better view can be obtained on which countries are net receiving or net sending. In addition, the total roaming volumes can be analysed as well. For example, a model can be fit to the total volumes of a certain country in a timespan (monthly, quarterly, yearly, etc.). It would be interesting to compare the forecast of the total volumes (forecasted a whole) with the sum of the separate forecasts of the in- and outbound volumes. In this way, it can be examined whether forecasting the in- and outbound volumes separately will give more accurate predictions compared to the forecast of the total roaming volumes as a whole.

Besides these general models, it could also be useful for operators to focus their attention on the peak volumes instead of the total yearly volumes or per subscriber use. The reason for this is because operators from net receiving countries will want to know if their network is able to handle the traffic spikes during a very busy period or a specific season where there are more tourists visiting the country. For example, Spain has the highest nights spent at tourist accommodation establishments in Europe [120] with 305.9 million nights spent in the country by non-residents. Spanish telecom operators will want to know during the summer holiday season how much incoming (peak) traffic there will be. So, building a model solely focused on peak use is beneficial for these countries as well. In BEREC's latest opinion report, it was observed that in the past year, end users complained about the quality of service (QoS) during roaming due to lower speeds while roaming in the EEA. This was, according to the same report, a result of some operators limiting their network speed during certain periods of the day. With this problem in mind and the expected continued increase in roaming use, these operators need to determine, based on the peak volume estimations, if their network infrastructure needs an upgrade to solve this problem and how to adapt their retail prices in order to recover these costs and remain competitive in the roaming market.

The models in the previous paragraphs fitted the roaming volumes with data from different potential influencing factors, however, it can also be useful to fit the roaming volumes without these factors by using techniques that only look at one dataset, i.e. the roaming use, and try to determine a function that fits this data in the best possible way. These are e.g. time series forecasting methods (ARIMA), non-linear regression, heuristics (genetic algorithm, simulated annealing, etc.), big data or machine learning techniques if large datasets are available. In that case, no analysis is made on the potential influencing factors, but these predictions can be compared to the ones obtained from the models discussed above to see whether predictions using multiple factors as input parameters are better than predictions using only the roaming volumes as input.

Once the above steps are performed, more accurate estimations are obtained from which abstracted regions can be defined and analysed. One possibility is to divide countries based on their wealth, e.g. Northern European countries which typically have a higher wealth per adult versus Eastern European countries which are typically less wealthy. Another way to group countries together could be based on whether it is a net sending or net receiving country. For this, the inbound volumes data needs to be obtained which was e.g. not possible in this dissertation. What could also be interesting is grouping countries together which have similar touristic activities during a certain season, e.g. countries that will have more tourists in the summer versus those with primarily winter tourists. Whether a certain way of defining abstracted regions is appropriate must be determined by analysing more reliable/larger datasets and the model predictions.

Up till now, the focus was on estimating the roaming volumes' evolution. However, to analyse the real impact on operators that is only the first step. The roaming volume estimates are used as input in the cost model of operators. This cost model is what is important for mobile operators. In such a model, different inputs, other than the roaming volume predictions, are used to estimate the total cost of providing roaming services for their subscribers; these include infrastructure costs, wholesale costs, other business operating costs, etc. From the output of this cost model, the actual change in costs (as a result of RLAH) can then be analysed. This is the real impact operators will experience. MNOs, MVNOs and cross-country operators have different costs as explained in sections 2.1.1 and 2.3.3, e.g. the wholesale prices between MNOs will typically be lower than the negotiated price between a MNO and MVNO. Single and cross-country operators will have a different amount of costs, e.g. a single country operator will have to pay wholesale prices to foreign operators in order to provide mobile roaming while a cross-country operator can avoid a part of these costs by steering traffic to its own network; however, their infrastructure costs will probably be higher than a single country operator. Consequently, each type of operator will have its own inputs for the cost model leading to different results. BEREC observed that MVNOs currently still lack negotiating power and therefore missing out on potential discounts on wholesale prices which are not pass on by the MNOs, the result is that the wholesale price for MVNOs are still close to the current price caps imposed by the EC and higher than the wholesale prices paid between MNOs. MNOs will typically have more infrastructure costs compared to MVNOs, because MVNOs will typically use an MNOs network infrastructure to provide roaming services for its customers[1]. Consequently, both operators will have a different cost model, it is therefore also useful to further differentiate the roaming volumes between MNOs and MVNOs. In this way, each type of operator can use their own estimated volumes as input in their respective cost model. It is from the outputs of the cost model that operators can analyse the actual impact on their network and the financial impact. This will allow operators to adapt their retail prices in order to recover potential investment costs in infrastructure upgrades or increases in the total wholesale prices they need to pay due to the increased roaming use of their own subscribers.

Lastly, since other parts in the world (Asia, South America, Africa, etc.) are also evolving into a roaming like at home principle (using the EEA as a benchmark) in their respective regions, it would be interesting to use the models with parameters and coefficients based on EEA countries to predict roaming volumes in other parts of the world to see whether the same accuracy can be achieved. Additionally, the usefulness of each input parameter can be tested and verified for other regions because it is possible that some parameters, which were valid for the EEA, are not applicable anymore to another region.

[1] There are exceptions possible. MVNOs can possess a certain amount of infrastructure. This affects the MVNOs' degree of dependence on the MNO.

Bibliography

[1] F. Lagaet, "Economic implications of reducing or abolishing roaming fees", Master's thesis, Ghent University, 2016-2017.

[2] J. Spruytte, M. Van der Wee, M. de Regt, S. Verbrugge, and D. Colle, "International roaming in the eu: Current overview, challenges, opportunities and solutions", *Telecommunications Policy*, vol. 41, no. 9, pp. 717–730, 2017.

[3] Investelecom, *Mobile virtual network operator*. [Online]. Available: `http://www.investele.com/documents/Mobile%20Virtual%20Network%20Operator.pdf` (visited on May 18, 2019).

[4] E. Sutherland, "International mobile roaming: Competition, economics and regulation", 2010.

[5] F. Bourassa, S. Paltridge, V. Weber, Y. Yokomori, and D. Ypsilanti, "Developments in international mobile roaming", 2016.

[6] ITU, *International mobile roaming strategic guidelines*, 2018. [Online]. Available: `https://www.itu.int/dms_pub/itu-d/opb/pref/D-PREF-EF.MIR03-2018-PDF-E.pdf` (visited on Feb. 23, 2019).

[7] European Commission, *Working document the initial findings of the sector inquiry into mobile roaming charges*, 2000. [Online]. Available: `http://ec.europa.eu/competition/sectors/telecommunications/archive/inquiries/roaming/working_document_on_initial_results.pdf` (visited on Feb. 23, 2019).

[8] E. Sutherland, "Intug europe report on gsm roaming prices-1999", *Available at SSRN 1547704*, 1999.

[9] C. Blackman and S. Forge, *Roaming: One Year After Implementation.* European Parliament, 2018.

[10] BEREC, *International roaming berec benchmark data report april 2017 - september 2017*, 2018. [Online]. Available: `https://berec.europa.eu/eng/document_register/subject_matter/berec/download/0/8011-international-roaming-berec-benchmark-da_0.pdf` (visited on Feb. 23, 2019).

[11] European Union, *Roaming in the eu*, May 14. [Online]. Available: https://europa.eu/youreurope/citizens/consumers/internet-telecoms/mobile-roaming-costs/index_en.htm (visited on May 18, 2019).

[12] M. Scotti, *Canada signs commitment to end data roaming charges, but it's not binding*, Global News, 2018. [Online]. Available: https://globalnews.ca/news/4096746/roaming-charges-canada-thing-of-the-past/ (visited on Feb. 25, 2019).

[13] CRASA Roaming Task Team, *The sadc home and away roaming project*, 2014. [Online]. Available: http://www.crasa.org/crasa-content/id/74/crasa-roaming-task-team/ (visited on Feb. 25, 2019).

[14] A. Gillwald and M. Mureithi, "Regulatory intervention or disruptive competition?: Lessons from east africa on the end of international mobile roaming charges", *Towards Evidence-based ICT Policy and Regulation policy paper series; 2010, v. 2, no. 1*, 2010.

[15] TeleGeography, *Denmark, china sign roaming agreement*, 2014. [Online]. Available: https://www.telegeography.com/products/commsupdate/articles/2014/10/15/denmark-china-sign-roaming-agreement/ (visited on Feb. 24, 2019).

[16] M. Raz-Chaimovich, *Overseas roaming reform in 6 months*, 2014. [Online]. Available: https://en.globes.co.il/en/article-overseas-roaming-reform-in-6-months-1000963374 (visited on Feb. 24, 2019).

[17] Australian Government, *Regulation impact statement: Trans-tasman mobile roaming public version*, 2013. [Online]. Available: https://ris.pmc.gov.au/sites/default/files/posts/2013/02/03-TTMR-RIS.pdf (visited on Feb. 23, 2019).

[18] A. Bender, *Australia and japan talk mobile roaming at first ict summit*, 2015. [Online]. Available: https://www.computerworld.com.au/article/566409/australia-japan-talk-mobile-roaming-first-ict-summit/ (visited on Feb. 24, 2019).

[19] BEREC, *International roaming berec benchmark data report april 2016 - september 2016*, 2017. [Online]. Available: https://berec.europa.eu/eng/document_register/subject_matter/berec/reports/7096-international-roaming-berec-benchmark-data-report-april-8211-september-2016 (visited on Feb. 23, 2019).

[20] BEREC, *International roaming berec benchmark data report october 2016 - march 2017*, 2017. [Online]. Available: https://berec.europa.eu/eng/document_register/subject_matter/berec/download/0/8011-international-roaming-berec-benchmark-da_0.pdf (visited on Feb. 23, 2019).

[21] BEREC, *International roaming berec benchmark data report october 2017 - march 2018*, 2018. [Online]. Available: `https://berec.europa.eu/eng/document_register/subject_matter/berec/reports/8251-international-roaming-berec-benchmark-data-report-october-2017-march-2018` (visited on Feb. 23, 2019).

[22] BEREC, *International roaming berec benchmark data report april 2018 - september 2018*, 2019. [Online]. Available: `https://berec.europa.eu/eng/document_register/subject_matter/berec/reports/8468-international-roaming-berec-benchmark-data-report-april-2018-september-2018` (visited on Apr. 16, 2019).

[23] BEREC, *Berec report on the wholesale roaming market*, 2016-02-12. [Online]. Available: `https://berec.europa.eu/eng/document_register/subject_matter/berec/reports/5745-berec-report-on-the-wholesale-roaming-market` (visited on Aug. 3, 2019).

[24] BEREC, *International roaming: Analysis of the impacts of "roam like at home" (rlah)*, 2014-12-17. [Online]. Available: `https://berec.europa.eu/eng/document_register/subject_matter/berec/opinions/4826-international-roaming-analysis-of-the-impacts-of-8220roam-like-at-home8221-rlah` (visited on Aug. 3, 2019).

[25] M. Falch and R. Tadayoni, "Regulation of international roaming data services within the eu", *International Economics and Economic Policy*, vol. 11, no. 1-2, pp. 81–95, 2014.

[26] BEREC, *Berec opinion on the functioning of the roaming market as input to ec evaluation*, 2019-06-13. [Online]. Available: `https://berec.europa.eu/eng/document_register/subject_matter/berec/opinions/8595-berec-opinion-on-the-functioning-of-the-roaming-market-as-input-to-ec-evaluation`.

[27] MVNO Europe, *Mvno europe is committed to the effective end of roaming charges*, 2015-10-27. [Online]. Available: `http://mvnoeurope.eu/mvnos-committed-to-effective-end-roaming-charges/` (visited on Aug. 3, 2019).

[28] Financial Times, *Definition of waterbed effect*. [Online]. Available: `http://lexicon.ft.com/Term?term=waterbed-effect` (visited on Feb. 23, 2019).

[29] BEREC, *Berec report on transparency and comparability of international roaming tariffs*, 2018. [Online]. Available: `https://berec.europa.eu/eng/document_register/subject_matter/berec/download/0/8312-berec-report-on-transparency-and-compara_0.pdf` (visited on Feb. 26, 2019).

[30] A. Lenhart, K. Purcell, A. Smith, and K. Zickuhr, "Social media & mobile internet use among teens and young adults. millennials.", *Pew internet & American life project*, 2010.

[31] T. Zhou, Y. Lu, and B. Wang, "Integrating ttf and utaut to explain mobile banking user adoption", *Computers in human behavior*, vol. 26, no. 4, pp. 760–767, 2010.

[32] V. Cisco, "Cisco visual networking index: Forecast and trends, 2017–2022", *White Paper*, 2018.

[33] T. J. Gerpott and S. Thomas, "Empirical research on mobile internet usage: A meta-analysis of the literature", *Telecommunications Policy*, vol. 38, no. 3, pp. 291–310, 2014.

[34] Biostat, *Why perform a meta-analysis?*, 2019. [Online]. Available: `https://www.meta-analysis.com/pages/why_do.php?cart=` (visited on May 4, 2019).

[35] J. E. Hunter and F. L. Schmidt, *Methods of meta-analysis: Correcting error and bias in research findings*. Sage, 2004.

[36] R. A. Peterson and S. P. Brown, "On the use of beta coefficients in meta-analysis", *Journal of Applied Psychology*, vol. 90, no. 1, p. 175, 2005.

[37] J. Boase and R. Ling, "Measuring mobile phone use: Self-report versus log data", *Journal of Computer-Mediated Communication*, vol. 18, no. 4, pp. 508–519, 2013.

[38] M. de Reuver, H. Bouwman, N. Heerschap, and H. Verkasalo, "Smartphone measurement: Do people use mobile applications as they say they do?", in *ICMB*, 2012, p. 2.

[39] A. Kivi, "Measuring mobile service usage: Methods and measurement points", *International Journal of Mobile Communications*, vol. 7, no. 4, pp. 415–435, 2009.

[40] R. Sharma, P. Yetton, and J. Crawford, "Estimating the effect of common method variance: The method—method pair technique with an illustration from tam research", *Mis Quarterly*, pp. 473–490, 2009.

[41] T. Smura, A. Kivi, and J. Töyli, "A framework for analysing the usage of mobile services", *info*, vol. 11, no. 4, pp. 53–67, 2009.

[42] F. D. Davis, R. P. Bagozzi, and P. R. Warshaw, "User acceptance of computer technology: A comparison of two theoretical models", *Management science*, vol. 35, no. 8, pp. 982–1003, 1989.

[43] I. Ajzen, "The theory of planned behavior", *Organizational behavior and human decision processes*, vol. 50, no. 2, pp. 179–211, 1991.

[44] E. M. Rogers, *Diffusion of innovations*. Simon and Schuster, 2010.

[45] M. Koufaris, "Applying the technology acceptance model and flow theory to online consumer behavior", *Information systems research*, vol. 13, no. 2, pp. 205–223, 2002.

[46] P. J. Hu, P. Y. Chau, O. R. L. Sheng, and K. Y. Tam, "Examining the technology acceptance model using physician acceptance of telemedicine technology", *Journal of management information systems*, vol. 16, no. 2, pp. 91–112, 1999.

[47] W. R. King and J. He, “A meta-analysis of the technology acceptance model”, *Information & management*, vol. 43, no. 6, pp. 740–755, 2006.

[48] M. Y. Chuttur, “Overview of the technology acceptance model: Origins, developments and future directions”, *Working Papers on Information Systems*, vol. 9, no. 37, pp. 9–37, 2009.

[49] S. Y. Yousafzai, G. R. Foxall, and J. G. Pallister, “Technology acceptance: A meta-analysis of the tam: Part 1”, *Journal of Modelling in Management*, vol. 2, no. 3, pp. 251–280, 2007.

[50] Y. Lee, K. A. Kozar, and K. R. Larsen, “The technology acceptance model: Past, present, and future”, *Communications of the Association for information systems*, vol. 12, no. 1, p. 50, 2003.

[51] F. F. Sniehotta, J. Presseau, and V. Araújo-Soares, *Time to retire the theory of planned behaviour*, 2014.

[52] F. Sniehotta, “An experimental test of the theory of planned behavior”, *Applied Psychology: Health and Well-Being*, vol. 1, no. 2, pp. 257–270, 2009.

[53] J.-H. Wu and S.-C. Wang, “What drives mobile commerce?: An empirical evaluation of the revised technology acceptance model”, *Information and management*, vol. 42, no. 5, pp. 719–729, 2005.

[54] V. Venkatesh, M. G. Morris, G. B. Davis, and F. D. Davis, “User acceptance of information technology: Toward a unified view”, *MIS quarterly*, pp. 425–478, 2003.

[55] J. C. Verhoeven, D. Heerwegh, and K. De Wit, “Information and communication technologies in the life of university freshmen: An analysis of change”, *Computers & Education*, vol. 55, no. 1, pp. 53–66, 2010.

[56] R. P. Bagozzi, “The legacy of the technology acceptance model and a proposal for a paradigm shift.”, *Journal of the association for information systems*, vol. 8, no. 4, p. 3, 2007.

[57] P. Nielsen and A. Fjuk, “The reality beyond the hype: Mobile internet is primarily an extension of pc-based internet”, *The Information Society*, vol. 26, no. 5, pp. 375–382, 2010.

[58] L. Armey, A. Vladar, and F. Pereira, “Are phones fun? usage and barriers to adoption for wireless data services among the younger generation of asians”, *International Journal of Electronic Business*, vol. 9, no. 1-2, pp. 8–25, 2011.

[59] M. Bina and G. M. Giaglis, “Perceived value and usage patterns of mobile data services: A cross-cultural study”, *Electronic Markets*, vol. 17, no. 4, pp. 241–252, 2007.

[60] M. Böhmer, B. Hecht, J. Schöning, A. Krüger, and G. Bauer, "Falling asleep with angry birds, facebook and kindle: A large scale study on mobile application usage", in *Proceedings of the 13th international conference on Human computer interaction with mobile devices and services*, ACM, 2011, pp. 47–56.

[61] H. Bouwman, C. Carlsson, C. Lopez-Nicolas, B. Mckenna, F Molina-Castillo, T. Tuunanen, and P. Walden, "Mobile travel services: The effect of moderating context factors", *Information Technology & Tourism*, vol. 13, no. 2, pp. 57–74, 2011.

[62] H. Bouwman, C. Carlsson, F. J. M. Castillo, G. M. Giaglis, and P. Walden, "Factors affecting the present and future use of mobile data services: Comparing the dutch, finnish and greek markets", *International Journal of Mobile Communications*, vol. 8, no. 4, pp. 430–450, 2010.

[63] I. D. Constantiou and M. Bina, "Exploring the use of mobile data services in europe: The cases of denmark and greece", in *Mobile Computing: Concepts, Methodologies, Tools, and Applications*, IGI Global, 2009, pp. 1296–1311.

[64] P. Harris, R. Rettie, and C. K. Cheung, "Adoption and usage of m-commerce: A cross-cultural comparison of hong kong and the united kingdom", *Journal of Electronic Commerce Research*, vol. 6, no. 3, pp. 210–224, 2005.

[65] A. Kaikkonen, "Full or tailored mobile web-where and how do people browse on their mobiles?", in *Proceedings of the International Conference on Mobile Technology, Applications, and Systems*, ACM, 2008, p. 28.

[66] R. A. Kalden, *Mobile internet traffic measurement and modeling based on data from commercial GPRS networks.* Mainz, 2004.

[67] R. Kalden, T. Varga, B. Wouters, and B. Sanders, "Wireless service usage and traffic characteristics in gprs networks", in *Teletraffic Science and Engineering*, vol. 5, Elsevier, 2003, pp. 981–990.

[68] S. Sarker and J. D. Wells, "Understanding mobile handheld device use and adoption", *Communications of the ACM*, vol. 46, no. 12, pp. 35–40, 2003.

[69] D. H. Shin, "A cross-national study of mobile internet services: A comparison of us and korean mobile internet users", in *International Enterprises and Global Information Technologies: Advancing Management Practices*, IGI Global, 2011, pp. 228–252.

[70] H. Verkasalo, "Empirical observations on the emergence of mobile multimedia services and applications in the us and europe", in *Proceedings of the 5th international conference on Mobile and ubiquitous multimedia*, ACM, 2006, p. 3.

[71] H. Verkasalo, "An international study of smartphone usage", *International Journal of Electronic Business*, vol. 9, no. 1-2, pp. 158–181, 2011.

[72] J. Yi, F. Maghoul, and J. Pedersen, “Deciphering mobile search patterns: A study of yahoo! mobile search queries”, in *Proceedings of the 17th international conference on World Wide Web*, ACM, 2008, pp. 257–266.

[73] I. Lee, B. Choi, J. Kim, and S.-J. Hong, “Culture-technology fit: Effects of cultural characteristics on the post-adoption beliefs of mobile internet users”, *International Journal of Electronic Commerce*, vol. 11, no. 4, pp. 11–51, 2007.

[74] I. Lee, J. Kim, B. Choi, and S.-J. Hong, “Measurement development for cultural characteristics of mobile internet users at the individual level”, *Computers in Human Behavior*, vol. 26, no. 6, pp. 1355–1368, 2010.

[75] L. Zhang, J. Zhu, and Q. Liu, “A meta-analysis of mobile commerce adoption and the moderating effect of culture”, *Computers in human behavior*, vol. 28, no. 5, pp. 1902–1911, 2012.

[76] K. O’Doherty, S. R. Hill, M. M. Mackay, and J. McPherson, “Mobile data service usage and preference: An investigation of australian consumers”, *International Journal of Mobile Communications*, vol. 8, no. 1, pp. 106–127, 2010.

[77] T. J. Gerpott, S. Thomas, and M. Weichert, “Personal characteristics and mobile internet use intensity of consumers with computer-centric communication devices: An exploratory empirical study of ipad and laptop users in germany”, *Telematics and Informatics*, vol. 30, no. 2, pp. 87–99, 2013.

[78] T. J. Gerpott, “Communication behaviors and perceptions of mobile internet adopters”, *info*, vol. 12, no. 4, pp. 54–73, 2010.

[79] X. He, P. P. Lee, L. Pan, C. He, and J. C. Lui, “A panoramic view of 3g data/control-plane traffic: Mobile device perspective”, in *International Conference on Research in Networking*, Springer, 2012, pp. 318–330.

[80] A. Kivi, “Diffusion of mobile data in finland”, *NETNOMICS: Economic Research and Electronic Networking*, vol. 11, no. 3, pp. 243–254, 2010.

[81] H. Liu, C.-N. Chuah, H. Zang, and S. Gatmir-Motahari, “Evolving landscape of cellular network traffic”, in *2012 21st International Conference on Computer Communications and Networks (ICCCN)*, IEEE, 2012, pp. 1–7.

[82] M. Z. Shafiq, L. Ji, A. X. Liu, and J. Wang, “Characterizing and modeling internet traffic dynamics of cellular devices”, *ACM SIGMETRICS Performance Evaluation Review*, vol. 39, no. 1, pp. 265–276, 2011.

[83] T. Smura, A. Kivi, and J. Toyli, “Mobile data services in finland: Usage of networks, devices, applications and content”, *International Journal of Electronic Business*, vol. 9, no. 1-2, pp. 138–157, 2011.

[84] Y. Jin, N. Duffield, A. Gerber, P. Haffner, W.-L. Hsu, G. Jacobson, S. Sen, S. Venkataraman, and Z.-L. Zhang, "Characterizing data usage patterns in a large cellular network", in *Proceedings of the 2012 ACM SIGCOMM workshop on Cellular networks: operations, challenges, and future design*, ACM, 2012, pp. 7–12.

[85] A. Hafsaoui, G. Urvoy-Keller, D. Collange, and M. Siekkinen, "A study of email usage and performance over cellular technology", in *The Second International Conference on Communications and Networking*, IEEE, 2010, pp. 1–8.

[86] M. Kamvar, M. Kellar, R. Patel, and Y. Xu, "Computers and iphones and mobile phones, oh my!: A logs-based comparison of search users on different devices", in *Proceedings of the 18th international conference on World wide web*, ACM, 2009, pp. 801–810.

[87] S. K. Kane, A. K. Karlson, B. R. Meyers, P. Johns, A. Jacobs, and G. Smith, "Exploring cross-device web use on pcs and mobile devices", in *IFIP Conference on Human-Computer Interaction*, Springer, 2009, pp. 722–735.

[88] A. Oulasvirta, T. Rattenbury, L. Ma, and E. Raita, "Habits make smartphone use more pervasive", *Personal and Ubiquitous Computing*, vol. 16, no. 1, pp. 105–114, 2012.

[89] P. Svoboda, F. Ricciato, R. Pilz, and E. Hasenleithner, *Composition of GPRS, UMTS traffic: snapshots from a live network.* na, 2006.

[90] P. Svoboda, *Measurement and modelling of Internet traffic over 2.5 and 3G cellular core networks.* na, 2008.

[91] H. Bouwman, C. Carlsson, P. Walden, and F. J. Molina-Castillo, "Trends in mobile services in finland 2004-2006: From ringtones to mobile internet", *info*, vol. 10, no. 2, pp. 75–93, 2008.

[92] T. J. Gerpott, S. Thomas, and M. Weichert, "Characteristics and mobile internet use intensity of consumers with different types of advanced handsets: An exploratory empirical study of iphone, android and other web-enabled mobile users in germany", *Telecommunications Policy*, vol. 37, no. 4-5, pp. 357–371, 2013.

[93] A. Ghose and S. P. Han, "An empirical analysis of user content generation and usage behavior on the mobile internet", *Management Science*, vol. 57, no. 9, pp. 1671–1691, 2011.

[94] H. Hämmäinen, A. Kivi, and H. Verkasalo, "Usage patterns of mobile subscribers: Regulatory observations in finland", in *16th Biennial ITS Conference*, 2006, pp. 12–16.

[95] S.-W. Lee, J.-S. Park, H.-S. Lee, and M.-S. Kim, "A study on smart-phone traffic analysis", in *2011 13th Asia-Pacific Network Operations and Management Symposium*, IEEE, 2011, pp. 1–7.

[96] A. Peslak, L.-J. Shannon, and W. Ceccucci, "An empirical study of cell phone and smartphone usage", *Issues in Information Systems*, vol. 12, no. 1, pp. 407–417, 2011.

[97] P. Sugai, "Exploring the impact of handset upgrades on mobile content and service usage", *International Journal of Mobile Communications*, vol. 5, no. 3, pp. 281–299, 2007.

[98] H. T. Verkasalo, "Handset-based measurement of smartphone service evolution in finland", *Journal of Targeting, Measurement and Analysis for Marketing*, vol. 16, no. 1, pp. 7–25, 2007.

[99] LAERD Statistics, *Moderator analysis with a dichotomous moderator using spss statistics*. [Online]. Available: `https://statistics.laerd.com/spss-tutorials/dichotomous-moderator-analysis-using-spss-statistics.php` (visited on May 4, 2019).

[100] European Commission, *Flash eurobarometer 454: The end of roaming charges within the eu*, 2017. [Online]. Available: `https://ec.europa.eu/commfrontoffice/publicopinion/index.cfm/ResultDoc/download/DocumentKy/79929` (visited on Mar. 7, 2019).

[101] J. Frost, *When should i use regression analysis?*, 2017. [Online]. Available: `https://statisticsbyjim.com/regression/when-use-regression-analysis/` (visited on Jul. 14, 2019).

[102] S. B. Green, "How many subjects does it take to do a regression analysis", *Multivariate behavioral research*, vol. 26, no. 3, pp. 499–510, 1991.

[103] D. C. Montgomery, *Introduction to statistical quality control.* John Wiley & Sons, 2007.

[104] Minitab, *Regression analysis: How do i interpret r-squared and assess the goodness-of-fit?*, May 2013. [Online]. Available: `https://blog.minitab.com/blog/adventures-in-statistics-2/regression-analysis-how-do-i-interpret-r-squared-and-assess-the-goodness-of-fit` (visited on May 18, 2019).

[105] Statistics Solutions, *Assumptions of linear regression.* [Online]. Available: `https://www.statisticssolutions.com/assumptions-of-linear-regression/` (visited on May 18, 2019).

[106] European Commission, *Eurostat*, 2019. [Online]. Available: `https://ec.europa.eu/eurostat/home` (visited on Mar. 7, 2019).

[107] European Commission, *Digital scoreboard - data & indicators*, 2018. [Online]. Available: https://digital-agenda-data.eu/ (visited on Mar. 7, 2019).

[108] European Commission, *Mobile broadband prices in europe 2017*, 2017. [Online]. Available: https://ec.europa.eu/digital-single-market/en/news/mobile-broadband-prices-europe-2017 (visited on Mar. 7, 2019).

[109] Credit Suisse, *Global wealth databook 2017*, Nov. 2017. [Online]. Available: https://www.credit-suisse.com/corporate/en/research/research-institute/global-wealth-report.html (visited on May 18, 2019).

[110] J. Tukey, *Exploratory Data Analysis*, ser. Addison-Wesley series in behavioral science. Addison-Wesley Publishing Company, 1977, ISBN: 9780201076165. [Online]. Available: https://books.google.be/books?id=UT9dAAAAIAAJ.

[111] M. Galarnyk, *Understanding boxplots*, 2018-09-12. [Online]. Available: https://towardsdatascience.com/understanding-boxplots-5e2df7bcbd51 (visited on Jul. 14, 2019).

[112] W.-C. Hong, Y. Dong, L.-Y. Chen, and C.-Y. Lai, "Taiwanese 3g mobile phone demand forecasting by svr with hybrid evolutionary algorithms", *Expert Systems with Applications*, vol. 37, no. 6, pp. 4452–4462, 2010.

[113] C. W. VanVoorhis and B. L. Morgan, "Understanding power and rules of thumb for determining sample sizes", *Tutorials in quantitative methods for psychology*, vol. 3, no. 2, pp. 43–50, 2007.

[114] S. S. Shapiro and M. B. Wilk, "An analysis of variance test for normality (complete samples)", *Biometrika*, vol. 52, no. 3/4, pp. 591–611, 1965.

[115] U. Paul, A. P. Subramanian, M. M. Buddhikot, and S. R. Das, "Understanding traffic dynamics in cellular data networks", in *2011 Proceedings IEEE INFOCOM*, IEEE, 2011, pp. 882–890.

[116] M. Z. Shafiq, L. Ji, A. X. Liu, J. Pang, and J. Wang, "Characterizing geospatial dynamics of application usage in a 3g cellular data network", in *2012 Proceedings IEEE INFOCOM*, IEEE, 2012, pp. 1341–1349.

[117] H. Tang, S. Xiao, G. Zhao, and C. Xing, "Traffic behaviors on the mobile internet", *Journal of Convergence Information Technology*, vol. 7, no. 7, 2012.

[118] H. Verkasalo, "Handset-based measurement of mobile service demand and value", *info*, vol. 10, no. 3, pp. 51–69, 2008.

[119] T. Yamakami, "Toward understanding the mobile internet user behavior: A methodology for user clustering with aging analysis", in *Proceedings of the Fourth International Conference on Parallel and Distributed Computing, Applications and Technologies*, IEEE, 2003, pp. 85–89.

[120] European Commission, *Tourism statistics*, 2018. [Online]. Available: https://ec.europa.eu/eurostat/statistics-explained/index.php/Tourism_statistics (visited on Jul. 30, 2019).

Appendix A

Collected data

The following appendices contain the collected data that was used in the regression analyses. For each variable, its original unit of measure as retrieved from the database and its relative value compared to the EU-28 average will be given.

A.1 Roaming data use

The roaming data use was retrieved from the International Roaming BEREC Benchmark Data Report [21] which contained data for Q4-2017. The collected data is summarised in Table A.1.

Table A.1: Roaming data use [21]

Country	Monthly average per subscriber in megabytes	Relative value
Austria	147,95	0,639
Belgium	104,81	0,453
Bulgaria	225,79	0,975
Croatia	171,07	0,739
Cyprus	688,11	2,972
Czech Republic	115,64	0,499
Denmark	369,32	1,595
Estonia	204,71	0,884
Finland	390,06	1,685
France	353,94	1,529
Germany	93,08	0,402
Greece	126,33	0,546
Hungary	187,60	0,810
Ireland	157,84	0,682
Italy	147,73	0,638
Latvia	232,12	1,003
Lithuania	137,13	0,592
Luxembourg	416,00	1,797
Malta	147,75	0,638
Netherlands	185,87	0,803
Poland	246,28	1,064
Portugal	126,93	0,548
Romania	298,62	1,290
Slovakia	83,57	0,361
Slovenia	218,35	0,943
Spain	248,51	1,073
Sweden	476,33	2,057
United Kingdom	180,98	0,782
EU-28 average	231,52	1,000

A.2 Domestic data use

The domestic data use was retrieved from the International Roaming BEREC Benchmark Data Report [21] which contained data for Q4-2017. The collected data is summarised in Table A.2.

Table A.2: Domestic data use [21]

Country	Monthly average per subscriber in gigabytes	Relative value
Austria	8,17	2,438
Belgium	0,84	0,250
Bulgaria	1,15	0,342
Croatia	2,02	0,602
Cyprus	2,24	0,667
Czech Republic	2,46	0,734
Denmark	5,09	1,518
Estonia	7,97	2,377
Finland	16,84	5,023
France	3,34	0,997
Germany	1,48	0,442
Greece	0,66	0,196
Hungary	1,29	0,386
Ireland	5,06	1,510
Italy	2,13	0,635
Latvia	5,45	1,626
Lithuania	4,16	1,239
Luxembourg	2,24	0,669
Malta	2,28	0,680
Netherlands	1,89	0,564
Poland	3,84	1,144
Portugal	1,26	0,375
Romania	1,63	0,486
Slovakia	0,69	0,205
Slovenia	2,07	0,619
Spain	2,10	0,625
Sweden	4,09	1,221
United Kingdom	1,45	0,433
EU-28 average	3,35	1,000

A.3 Travelling in EU-28

The travel frequency was retrieved from the Eurobarometer survey [100]. These were collected shortly after the introduction of RLAH. The collected data is summarised in Table A.3.

Table A.3: Travelling in EU-28 [100]

Country	% of respondents who travelled to an EU-28 country	Relative value
Austria	66	1,737
Belgium	61	1,605
Bulgaria	17	0,447
Croatia	27	0,711
Cyprus	35	0,921
Czech Republic	40	1,053
Denmark	54	1,421
Estonia	34	0,895
Finland	40	1,053
France	37	0,974
Germany	50	1,316
Greece	12	0,316
Hungary	27	0,711
Ireland	24	0,632
Italy	24	0,632
Latvia	33	0,868
Lithuania	31	0,816
Luxembourg	85	2,237
Malta	33	0,868
Netherlands	54	1,421
Poland	24	0,632
Portugal	20	0,526
Romania	21	0,553
Slovakia	46	1,211
Slovenia	66	1,737
Spain	20	0,526
Sweden	50	1,316
United Kingdom	41	1,079
EU-28 average	38	1,000

A.4 Wealth per adult

The wealth per adult was collected from the 2017 Credit Suisse Research Institute's Global Wealth Databook [109]. The collected data is summarised in Table A.4.

Table A.4: Wealth per adult [109]

Country	US dollar	Relative value
Austria	231,37	1,534
Belgium	313,05	2,075
Bulgaria	23,98	0,159
Croatia	35,95	0,238
Cyprus	100,31	0,665
Czech Republic	61,49	0,408
Denmark	286,71	1,901
Estonia	57,81	0,383
Finland	161,06	1,068
France	280,58	1,860
Germany	214,89	1,425
Greece	108,13	0,717
Hungary	37,59	0,249
Ireland	232,95	1,544
Italy	217,79	1,444
Latvia	33,96	0,225
Lithuania	24,60	0,163
Luxembourg	412,13	2,732
Malta	140,63	0,932
Netherlands	253,21	1,679
Poland	31,79	0,211
Portugal	109,36	0,725
Romania	20,32	0,135
Slovakia	34,78	0,231
Slovenia	79,10	0,524
Spain	191,18	1,267
Sweden	249,77	1,656
United Kingdom	279,05	1,850
EU-28 average	150,84	1,000

A.5 Usefulness

The usefulness was retrieved from the Eurobarometer survey [100]. These were collected shortly after the introduction of RLAH. The collected data is summarised in Table A.5.

Table A.5: Usefulness [100]

Country	% respondents who thinks RLAH is beneficial for them	Relative value
Austria	81	1,099
Belgium	72	0,977
Bulgaria	74	1,004
Croatia	77	1,045
Cyprus	79	1,072
Czech Republic	67	0,909
Denmark	82	1,112
Estonia	70	0,950
Finland	78	1,058
France	59	0,800
Germany	79	1,072
Greece	61	0,828
Hungary	81	1,099
Ireland	91	1,234
Italy	67	0,909
Latvia	63	0,855
Lithuania	68	0,922
Luxembourg	77	1,045
Malta	83	1,126
Netherlands	78	1,058
Poland	73	0,990
Portugal	67	0,909
Romania	60	0,814
Slovakia	65	0,882
Slovenia	74	1,004
Spain	79	1,072
Sweden	83	1,126
United Kingdom	76	1,031
EU-28 average	73,7	1,000

A.6 Turn off mobile data

The turn off mobile data was retrieved from the Eurobarometer survey [100]. These were collected shortly after the introduction of RLAH. The collected data is summarised in Table A.6.

Table A.6: Turn off mobile data [100]

Country	% respondents turning off mobile data when travelling in EU-28	Relative value
Austria	40	1,432
Belgium	34	1,217
Bulgaria	20	0,716
Croatia	41	1,468
Cyprus	12	0,430
Czech Republic	26	0,931
Denmark	36	1,289
Estonia	26	0,931
Finland	37	1,325
France	32	1,146
Germany	33	1,182
Greece	18	0,645
Hungary	22	0,788
Ireland	45	1,611
Italy	25	0,895
Latvia	24	0,859
Lithuania	21	0,752
Luxembourg	28	1,003
Malta	11	0,394
Netherlands	37	1,325
Poland	16	0,573
Portugal	13	0,465
Romania	20	0,716
Slovakia	20	0,716
Slovenia	18	0,645
Spain	38	1,361
Sweden	39	1,396
United Kingdom	50	1,790
EU-28 average	27,9	1,000

A.7 Mobile penetration

The mobile penetration was retrieved from the Digital Scoreboard database of the European Commission [107]. The collected data is summarised in Table A.6.

Table A.7: Mobile penetration [107]

Country	Number of SIM cards per 100 people	Relative value
Austria	87,7	0,909
Belgium	75,7	0,784
Bulgaria	91,4	0,946
Croatia	84,0	0,870
Cyprus	105,8	1,096
Czech Republic	82,3	0,853
Denmark	129,0	1,337
Estonia	132,7	1,375
Finland	154,3	1,599
France	89,8	0,931
Germany	79,4	0,822
Greece	65,7	0,681
Hungary	48,7	0,505
Ireland	101,5	1,051
Italy	86,1	0,892
Latvia	120,0	1,243
Lithuania	82,4	0,853
Luxembourg	131,7	1,365
Malta	95,9	0,994
Netherlands	90,5	0,938
Poland	154,9	1,605
Portugal	69,0	0,715
Romania	85,4	0,885
Slovakia	83,6	0,866
Slovenia	70,5	0,730
Spain	94,2	0,976
Sweden	121,5	1,259
United Kingdom	88,6	0,918
EU-28 average	96,5	1,000

A.8 Retail price

The retail price was retrieved from the 2017 Mobile Broadband Prices in Europe report [108]. The collected data is summarised in Table A.8.

Table A.8: Retail price [108]

Country	Relative number between 1-4	Relative value
Austria	1	0,459
Belgium	3	1,377
Bulgaria	2	0,918
Croatia	2	0,918
Cyprus	4	1,836
Czech Republic	4	1,836
Denmark	2	0,918
Estonia	1	0,459
Finland	3	1,377
France	1	0,459
Germany	2	0,918
Greece	4	1,836
Hungary	4	1,836
Ireland	3	1,377
Italy	1	0,459
Latvia	1	0,459
Lithuania	1	0,459
Luxembourg	1	0,459
Malta	3	1,377
Netherlands	3	1,377
Poland	1	0,459
Portugal	3	1,377
Romania	2	0,918
Slovakia	3	1,377
Slovenia	1	0,459
Spain	2	0,918
Sweden	1	0,459
United Kingdom	2	0,918
EU-28 average	2,2	1,000

A.9 RLAH awareness

The RLAH awareness was retrieved from the Eurobarometer survey [100]. These were collected shortly after the introduction of RLAH. The collected data is summarised in Table A.9.

Table A.9: RLAH awareness [100]

Country	% respondents who is aware of RLAH	Relative value
Austria	85	1,129
Belgium	82	1,089
Bulgaria	73	0,969
Croatia	82	1,089
Cyprus	74	0,983
Czech Republic	84	1,116
Denmark	81	1,076
Estonia	82	1,089
Finland	72	0,956
France	59	0,784
Germany	82	1,089
Greece	53	0,704
Hungary	78	1,036
Ireland	74	0,983
Italy	67	0,890
Latvia	77	1,023
Lithuania	68	0,903
Luxembourg	87	1,155
Malta	77	1,023
Netherlands	85	1,129
Poland	79	1,049
Portugal	66	0,876
Romania	63	0,837
Slovakia	78	1,036
Slovenia	86	1,142
Spain	68	0,903
Sweden	83	1,102
United Kingdom	63	0,837
EU-28 average	75,3	1,000