TECHNISCHE UNIVERSITÄT WIEN
Vienna University of Technology

DIPLOMARBEIT

Airport passenger traffic forecasting

ausgeführt zum Zwecke der Erlangung des Akademischen Grades eines
Diplom-Ingenieurs

unter der Leitung von
Univ.Prof. Mag.rer.soc.oec. Dr.rer.soc.oec. Walter S.A. SCHWAIGER, MBA
E330
Institut für Managementwissenschaften

eingereicht an der Technischen Universität Wien
Fakultät für Maschinenwesen und Betriebswissenschaften

von

Todor Kucarov

0927539

Gassergasse 27/9
A-1050 Wien

Wien, am 13. März 2014

Ich habe zur Kenntnis genommen, dass ich zur Drucklegung meiner Arbeit unter der Bezeichnung

DIPLOMARBEIT

nur mit Bewilligung der Prüfungskommission berechtigt bin.

Ich erkläre weiters an Eides statt, dass ich meine Diplomarbeit nach den anerkannten Grundsätzen für wissenschaftliche Abhandlungen selbstständig ausgeführt habe und alle verwendeten Hilfsmittel, insbesondere die zugrunde gelegte Literatur genannt habe.

Weiters erkläre ich, dass ich dieses Diplomarbeitsthema bisher weder im In- noch im Ausland (einer Beurteilerin/ einem Beurteiler zur Begutachtung) in irgendeiner Form als Prüfungsarbeit vorgelegt habe und dass diese Arbeit mit der vom Begutachter beurteilten Arbeit übereinstimmt

Wien, am 13. März 2014

(Todor Kucarov)

Danksagung

An dieser Stelle möchte ich mich herzlich bei allen bedanken, die mich bei dieser Diplomarbeit tatkräftig unterstützt haben.

Einen ganz besonderen herzlichen Dank an meine ganze Familie, vor allem meinen Eltern *Aneta* und *Krasimir*, ohne deren Unterstützung mein Studium nicht möglich gewesen wäre. Ihnen möchte ich diese Diplomarbeit widmen.

Großer Dank gilt auch an meine Freundin *Mila* und an meine in Österreich gewonnen Freunde *Matthias Ondra*, *Clemens Leidl* und *Georg Katzlinger* die mich während meines zeitintensiven Studiums stets moralisch unterstützt und motiviert haben.

Ich möchte mich bei Herrn *Univ. Prof. Mag. Dr. Walter S.A. Schwaiger* für die hervorragende Betreuung und Hilfsstellung bei der Durchführung dieser Diplomarbeit bedanken. Schon bei den Grundlagenvorlesungen weckte Herr Schwaiger mein Interesse am Gebiet der Finanzwirtschaft und dem Risikomanagement und stand mir in den vertiefenden Lehrveranstaltungen sowie bei der Diplomarbeit immer mit Rat und Tat zur Seite.

Ebenso möchte ich mich auch bei allen Mitarbeiter der Firma *prevero* bedanken, die mir sowohl durch fachliche Unterstützung als auch durch immerwährende organisatorische Hilfestellung das Verfassen dieser Arbeit erst ermöglichten.

Wien, am 13. März 2014 Todor Kucarov

Abstract

The air passengers traffic forecasting has an enormous influence on the development of airport strategies and master plans with respect to both the airside and the landside. As the airport managers take their decisions based also on expected passenger volumes it is very important for them to be able to estimate the future demands as accurate as possible. This work deals with exactly this problem- how to develop a model to forecast air passenger demands based only on the historical passenger numbers data.

The work will give a detailed description of the theoretical fundaments and the process of construction, calibration and validation for two methods that can be adopted by an airport- the Rational passenger numbers planning and the Box-Jenkins ARIMA methods. By making a forecast for the Vienna International Airport next to the theoretical description of the techniques also their correct use in practise will be shown and their accuracy will be proven.

Table of Contents

1 Introduction

Problem Statement

For many businesses and industries forecasting is a crucial part of the business strategy development and has a central role in the planning process. This is even more emphasised in businesses, such as the air transport, that are very expensive and where careful planning can allow the management to cut expenses and reduce risk.

This work will look at two forecasting strategies that can be adopted by an airport and which make their predictions based on the historical passenger demand and actual flights. The process of construction, calibration and validation for both models will be explained in detail and using those techniques a prediction of the passengers demand at the Vienna International Airport will be made.

Expected results

By reading this work the reader should get familiar with the theoretical fundamentals and the process of construction, calibration and validation of the two prediction models described: the rational passenger numbers planning and the Box-Jenkins ARIMA Methodology. After the reader is made familiar with the theory of the distinct technique, a specific numerical example should be given in order to make the presentation also intuitively accessible. By using both methods in a forecasting procedure for the Vienna International Airport this work shows how the techniques are to be used in practice and which are the problematic points that may lead to either inaccurate estimates of the passenger numbers or to leave misleading feelings of security in the analyst.

Methodology

This work shows two models for predicting airport passenger demands. It starts with a definition and an explanation of the rational planning and the multi-period Gaussian enterprise (portfolio) framework along with their main characteristics. After the reader is made familiar with the theoretical and mathematical foundations behind the rational passenger numbers planning the fraction based Gaussian enterprise model is constructed and calibrated in order to calculate the multi-period planned passenger numbers trajectory for the Vienna International Airport. As forecasts could never be 100% accurate the work continues with the explanation and calculation of the multi-period passenger numbers quantiles and the 90% uncertainty corridor. Than in order to exam how good the model performs a model validation with an out of sample test is made. The section ends with the process of subsequent planning and what kind of check and act activities can be made in order to make sure that the prediction are achieved as precise as possible.

The second section is dedicated to the ARIMA models. After a description of this Box-Jenkins methodology and the method of forecasting among with all the statistical terms that are needed in order to understand it are characterised, a passenger demand forecast of the passenger volumes at the Vienna International Airport using ARIMA is made. The prediction includes a comprehensive analysis of the historical data and a detailed description of every step made through the process of forecasting. This is followed by an out of sample test for three different periods of time.

Finally, a conclusion summarizes the main findings of the work and the key points that have been discussed chronologically.

2 Rational Passenger Numbers Planning

The "rational passenger numbers planning" relates to the rational expectation theory developed by Lucas (1978), where the rational expectations are: forward looking, related to an uncertain economy and calculated according to a probabilistic model. When applied to the passenger numbers planning problem the rational expectations are computed according to the probabilistic passenger numbers model, which on the other hand is constructed within a Gaussian management framework. This framework was used by Black and Scholes (1973) and Merton (1973) by the developing of the option pricing theory and the intertemporal portfolio optimization theory.

In the rational passenger numbers planning process the passenger numbers yearly realisation is taken and translated according to the fraction-based segment passenger numbers model into a planned passenger numbers trajectory over the twelve months of the year- the twelve monthly sub-periods. The trajectory provides the passenger numbers prediction up to the yearly passenger numbers. The model allows next to the calculation of the trajectory also the computation of uncertainty corridors, which are accompanying it. As the model is stated in the conditional notation it can be used not only for the initial planning at the beginning of the planning horizon but also for the subsequent planning over the different sub-periods. By calculating rolling forecasts and rolling passenger numbers volatilities one can use a risk-based performance management system to make sure that the annual passenger numbers are achieved over time.

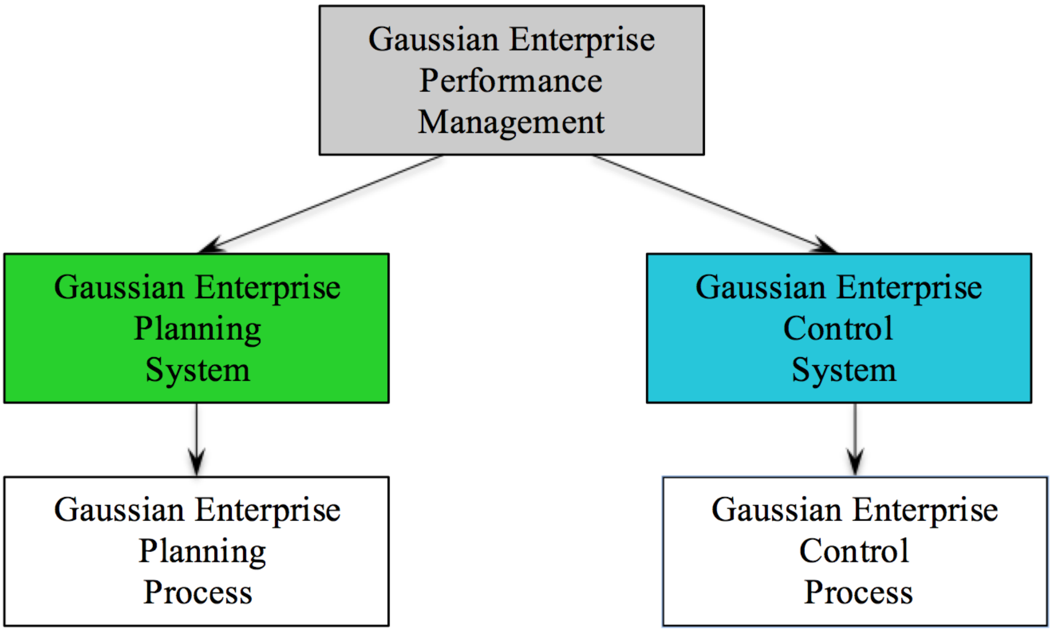
2.1 The rational passenger numbers planning technique

The "Rational passenger numbers planning" corresponds to the concept of rational expectations introduced by Lucas (1978). This is a theory, where the rational expectations are forward looking, related to an uncertain economy and calculated according to a probabilistic model, which, based on a historical data, calculates how big the chances of an event to happen again are.

When applied to the passenger numbers planning problem the rational passenger numbers expectations are the rational passenger numbers plans for the different planning horizons. These plans are calculated according to the probabilistic passenger numbers model, which stochastic nature originates from the probabilistic information structure that underlines the stochastic business environment. According to the probabilistic foundation the rational passenger number plans consist out of the passenger numbers forecast and the uncertainties surrounding the forecast- the passenger numbers volatility.

The rational passenger number planning is based upon the probabilistic passenger numbers model, which is constructed within a multi-period Gaussian enterprise (portfolio) framework, which as the name says incorporates a multi-period portfolio context, so that not only time diversification effects but also portfolio diversification effects are considered. This management framework was introduced by Black/- Scholes (1973) and Merton (1973) in the option pricing theory as well as in the intertemporal portfolio management theory. The multi-period Gaussian enterprise (portfolio) framework decomposes in to Gaussian enterprise planning system, where by using the fraction-based passenger numbers model rational expectations over the future planned passenger numbers and the related risks for the different segments are formed and Gaussian enterprise control system, where over time intermediate Check- and Act- activities are executed to ensure the realization of the annual passenger numbers.

Figure 1: Gaussian Enterprise Performance Management



Source: Schwaiger, Bös and Kronfellner 2012 P.54

For passenger numbers performance management purposes a discrete time approach is adequate as the control strategies relate to the passenger numbers performances over longer time horizons (e.g. months). Therefore the probabilistic passenger numbers model is built in a discrete time management framework. A discrete time Gaussian stochastic process is used for the modelling of the probabilistic information structure of the future passenger numbers developments. This method consists out of twelve Gaussian random variables (a term explained in the next paragraph) for every one of the portfolio segments. These specify the probabilistic passenger numbers segments volumes in the twelve months of the year and linear functions on this normally distributed variables define the monthly and the annual passenger numbers. As shown in the next paragraph due to this linearity of the aggregation functions the monthly and the annual passenger numbers are normally distributed as well. The forecasts correspond to the expected mean values and the volatilities are used for the calculation of the uncertainty bounds that are associated with the passenger numbers forecast in the stochastic business environment.

The Gaussian (or normal) distribution belongs to one of the theoretically and practically most important probability distributions due to the fact that it can be used for the description of many real situations. Mathematically: every random variable X that has the density

$$f(x) = \frac{1}{2\sqrt{\pi}} e^{-\frac{x^2}{2}} \tag{1}$$

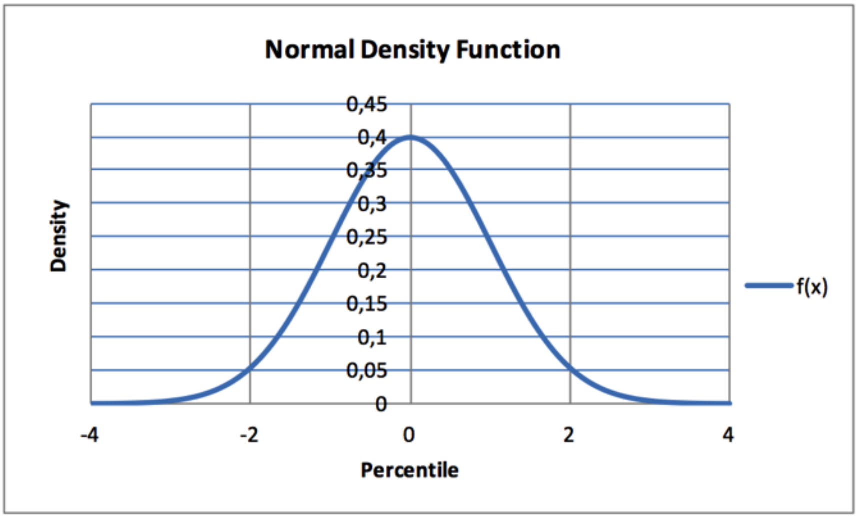
is called normally distributed.

We define:

$$X \sim N(0; 1). \tag{2}$$

In the next Figure 2 one can see the graph of the density of the N(0, 1)

Figure 2: Normal Density Function



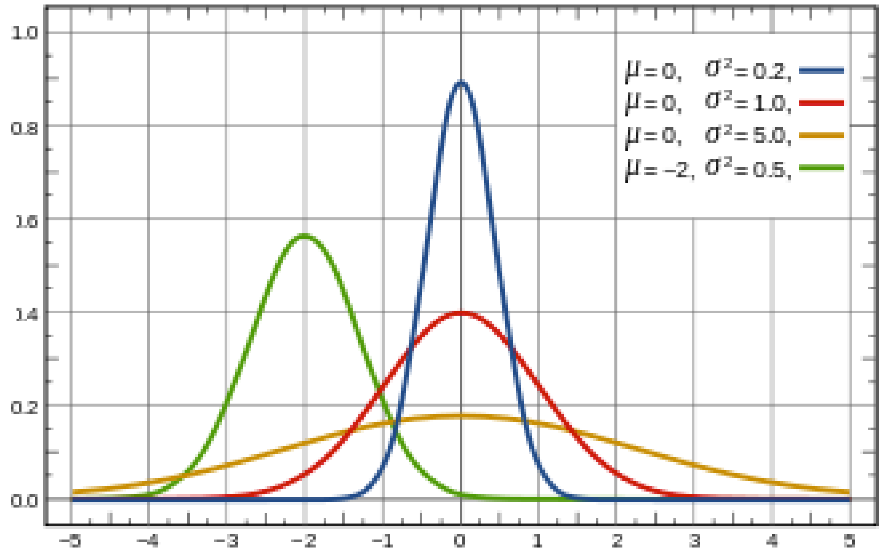
The area under the curve equals one due to the constant of integration $1/(2\sqrt{\pi})$. Please note that all distributions, which emerge from the normal distribution through a linear transformation, are also normally distributed. These build the family of the normal distributions. Mathematically this means that if $X \sim N(0; 1)$ and X is linearly transformed to $Y = \mu + \sigma X$ than Y has the expected value of μ, the variance σ^2 and the density:

$$f_Y(y) = \frac{1}{\sqrt{2 \cdot \pi \cdot \sigma^2}} \cdot \exp\left(-\frac{1}{2} \cdot \left(\frac{y-\mu}{\sigma^2}\right)^2\right). \tag{3}$$

One can say that Y is normally distributed: $Y \sim N(\mu; \sigma^2)$ (*Arens, Hettlich, Karpfinger, Kockelhorn, Lichtenberg and Stachel, 2008 P.1348-1350).*

Figure 3 on the next page graphically represents how the bell curve is changing when μ respectively σ are changed. One can see that when the mean of the random variable is changed, than the curve shifts to the left or right depending on the means value. On the other hand when the variance is changed, than the curve gets flatter if σ converges towards zero and is or steeper if σ tends toward infinity *(Arens, Hettlich, Karpfinger, Kockelhorn, Lichtenberg and Stachel, 2008 P.1348- 1350).*

Figure 3: Normal Density Function for different means and volatilities



Source: http://en.wikipedia.org/wiki/File:Normal_Distribution_PDF.svg

The fraction based Gaussian enterprise model used in this work enables the computation of passenger numbers predictions and passenger numbers volatilities for different time horizons up to one year. The planned passenger numbers trajectory consists out of the accumulated monthly passenger numbers forecasts and the monthly passenger numbers volatilities, which are used to compute the prediction uncertainty corridors that are associated with the passenger numbers forecasts for the different planning horizons. In the subsequent passenger numbers planning, described in the last section of this chapter, the Gaussian model is used to compute the rolling passenger numbers predictions and the rolling passenger numbers volatility.

In the following two parts the planned passenger numbers will be calculated. These contain out of the accumulated monthly passenger numbers forecast and the multi-period 90% enterprise passenger numbers uncertainty corridor around the forecasts. The upcoming sections will start with a theoretical part, where the model construction and the theory behind the technique is explained and will continue with the calibration process for which the historical data from the Vienna International Airport will be used. In order to make the calculation also intuitively accessible a concrete numerical examples are given for every step.

2.2 Multi-period planned passenger numbers trajectory

The multi-period Gaussian enterprise management framework uses the fraction-based approach, where the annual passenger numbers are allocated to the twelve months of the year according to the monthly fractions of the annual figures. This fraction based passenger numbers allocation gives the planned traffic trajectory over the twelve months and the means of the Gaussian random variables are used for the path modelling.

2.2.1 Construction of the Multi-period planned passenger numbers trajectory

In the fraction based Gaussian enterprise model the annual passenger numbers are defined as the sum of the twelve monthly fraction passenger numbers, which on their side are equal to the sum of the monthly fraction segment passenger numbers, a mathematical definition shown in equation (4). Here the index P stand for passengers, (k) stands for the different segments, (PF) stands for portfolio, as these values refer to the whole enterprise portfolio and not just single elements, M is the single-period notation and (s_0) signals that the multi-period view refers to the time slightly before month 1 begins.

$$
\begin{aligned}
&\tilde{x}_{P(PF)}(s_0) = \sum_{M=1}^{12} \tilde{x}_{P(PF),M}(s_0) \\
&\textit{with} \\
&\tilde{x}_{P(PF),M}(s_0) = \sum_{k} \tilde{x}_{P(k),M}(\mathrm{s}_0) = \sum_{k} x_{P(k)}^{Plan} . \tilde{a}_{P(k),M}(\mathrm{s}_0) \\
&\textit{and} \\
&\tilde{a}_{P(k),M}(\mathrm{s}_0) \sim N\left(\mu_{a(k),M}(\mathrm{s}_0); \sigma_{a(k),M}^2(\mathrm{s}_0)\right) \quad \forall M \\
&\textit{such that} \\
&\tilde{x}_{P(PF)}(s_0) = \sum_{M=1}^{12} \sum_{k} x_{P(k)}^{PLAN} . \tilde{a}_{P(k),M}(\mathrm{s}_0) .
\end{aligned}
\tag{4}
$$

The last line of equation (4) shows the fraction-based Gaussian enterprise model, where the planned segment passenger numbers $x_{P(k)}^{PLAN}$ are multiplied with the passenger numbers segment fractions $a_{P(k),M}$ and summed up for the different segments and over the twelve months of the year. As the initial monthly segment passenger numbers $\tilde{x}_{P(k),M}(s_0)$, the initial monthly portfolio passenger numbers $\tilde{x}_{P(PF),M}(s_0)$ and the initial annual portfolio passenger numbers $\tilde{x}_{P(PF)}(s_0)$ are linear functions of the normally distributed passenger numbers segment fractions $a_{P(k),M}$, one can say that these are normally distributed as well, something that was shown in the previous section.

The s_0- conditional initial passenger numbers forecast for the annual passenger numbers is shown in equation (5) in form of the conditional annual segment passenger numbers expectations $E[\tilde{x}_{P(PF)}|s_0]$

$$E[\tilde{x}_{P(PF)}|s_0] = \sum_{M=1}^{12} \sum_{k} x_{P(k)}^{PLAN} . E\left[\tilde{a}_{P(k),M}\middle|\mathrm{s}_0\right] \tag{5}$$

In the fraction-based Gaussian enterprise model the location parameters μ and the dispersion parameters σ are directly linked to the moments (to the expected values) of the distribution. This means that in the used model the expected mean value of the annual passenger numbers $\mu_{P(PF)}(s_0)$ is the mathematical representation of the passenger numbers forecast over the next year and the mean $\mu_{a(k),M}$ is equal to the expected fraction based segment passenger numbers. Equation (5) can be re-written as follows:

$$\begin{aligned} \mu_{P(PF)}(s_0) &= \sum_{M=1}^{12} \sum_{k} x_{P(k)}^{PLAN} . \mu_{a(k),M}(s_0) \\ &= x_{P(PF)}^{PLAN} \cdot \sum_{M=1}^{12} \mu_{a,M}(s_0) \\ &= x_{P(PF)}^{PLAN} \cdot \mu_a(s_0) \end{aligned} \tag{6}$$

In equation (6) the expected enterprise passenger numbers mean is equal to the location parameter of the passenger numbers normal distribution. This equality is used to term the passenger numbers forecast also as the enterprise passenger numbers mean parameter $\mu_{P(PF)}$.

Until now the focus was on the monthly and annual passenger numbers. In order to calculate the multi-period planned passenger numbers trajectory one has to take an extension by including also intermediate planning horizons that vary between one month and one year. This is done in order to calculate the passenger number means over multiple sub-periods. The sequence of all intermediating passenger numbers means from one month to one year gives the multi-period planned passenger numbers trajectory. The mathematical expression is given in the following equation (7) where T is the multi-period notation of the twelve months of the year (T = 1, 2, 3, … , 12):

$$\mu_{P(PF),T}(s_0) = \sum_{M \leq T} \sum_{k} x_{P(k)}^{PLAN} \cdot \mu_{a(k),M}(s_0) \quad (7)$$

As one can see the first value of the multi-period planned passenger trajectory is equal to the forecasted value of month one. For the multi-period time horizons the monthly passenger numbers are added to give the intermediate forecast. The annual prediction is equal to the accumulated forecast for the twelve periods.

Being familiar with the construction of the fraction-based Gaussian enterprise model the next part of this work will show the process of it`s calibration, where the multi-period planned passenger numbers trajectory for the Vienna International Airport will be calculated by using historical data. In order to give the reader a better feeling of how the model is used for every step of the calculation a numerical example will be given.

2.2.2 Historical Calibration of the Multi-period planned passenger numbers trajectory

As shown above the monthly segment passenger numbers parameters $\mu_{a(k),M}$ are equal to the location parameters of the normally distributed monthly segment passenger numbers fractions. These parameters are calculated by averaging the realized monthly segment passenger numbers over the last years. This expression is shown mathematically in the next equation (8), where Y represents the number of years used in the estimation and (y) shows the year of the realization:

$$\mu_{a(k),M} = \sum_{y(ear)=1}^{Y} \frac{a_{P(k),M(y)}}{Y} \quad (8)$$

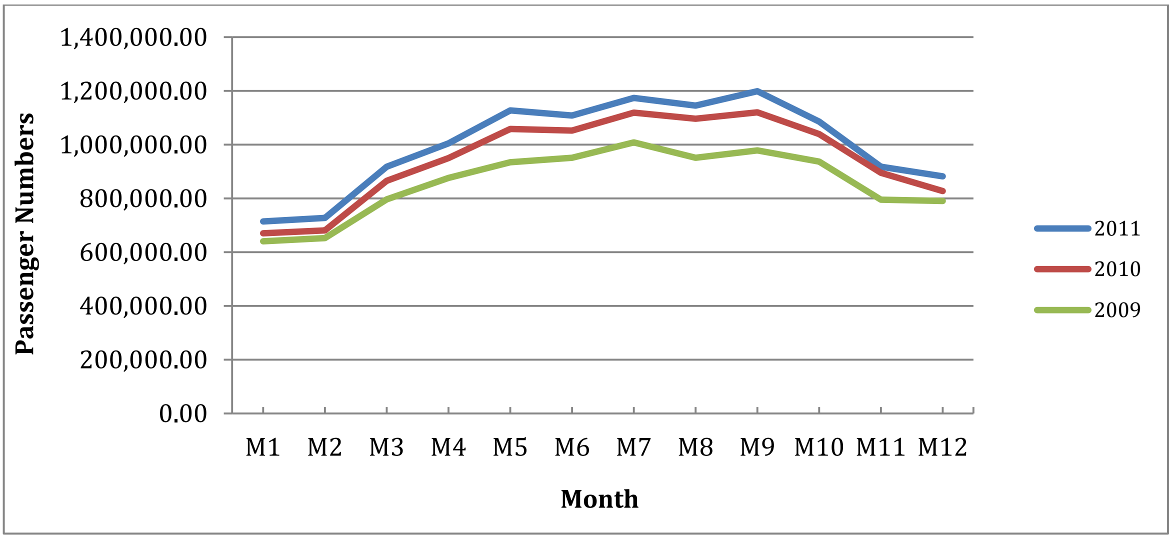
As the fraction-based Gaussian enterprise model consists of more then one realized monthly segment passenger numbers fractions, these have to be derived as well. This is achieved by dividing the monthly passenger numbers by the annual figures:

$$a_{P(k),M(y)} = \frac{x_{P(k),M(y)}}{x_{P(k)(y)}} \quad (9)$$

To set up the fraction-based Gaussian enterprise performance management a plan for the passenger numbers for the year 2012 should be made. In order to do this one has to take a closer look at the historical data and to analyse it deeply. For this first prediction the passengers will be divided in three subgroups- Passengers traveling from/to EU Schengen (S), Passengers travelling from/to EU Non-Schengen (NS) and Passengers travelling from/to Rest of the World (ROW) destinations. The model will be constructed by using the years 2009, 2010, 2011, which means that the forecast for 2012 will be made without knowing the historical numbers for 2012 and then in the validation part it will be compared to the real numbers in order to get feedback about the accuracy of the used methodology.

Before making a prediction for the separate months a prediction for the whole year for every one of the three segments is made. By plotting the data for the years 2009, 2010 and 2011, which can be seen in Figure 4, and examining it, it became clear that a good approximation for the growth of the annual passenger numbers from the first group (Schengen) is to take 5.5%.

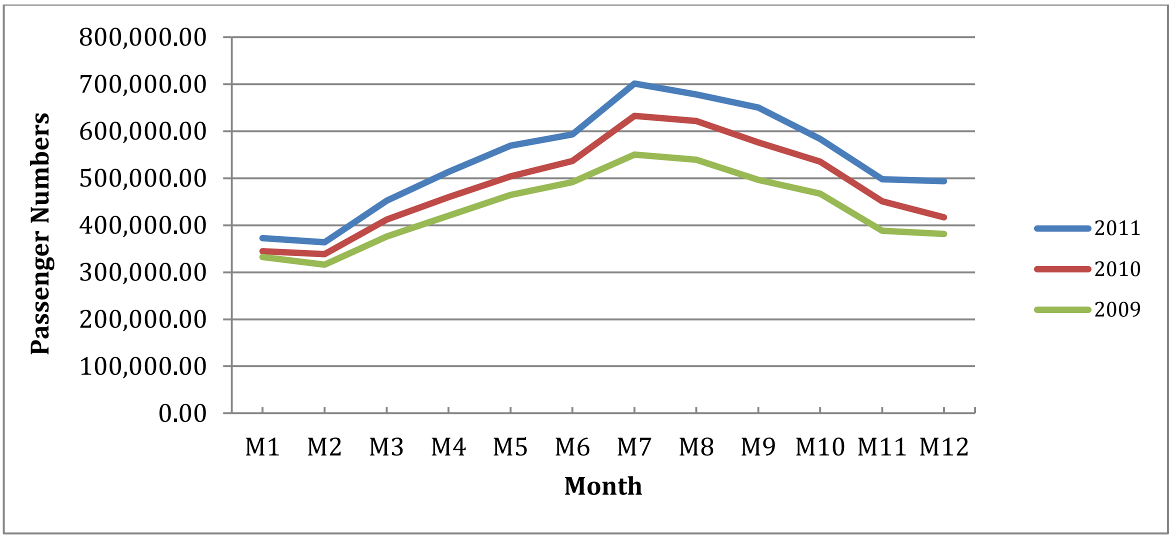
Figure 4: Historical data- EU Schengen passengers



This number is chosen because it can be believed that the trend of lowering the passenger numbers growth from the last two years 2010 and 2011 will continue, as the economy in the Schengen countries gets more stable after the crisis in 2008 and there will be no more big growths in the passenger demand like between 2009 and 2010.

For the second group (EU Non-Schengen) a good approximation will be to take the average percentage growth for the years 2010 and 2011, which is 11.28% and to use it as a growth for the year 2012. This percentage is chosen because, as it can be seen in Figure 5 during the crisis there were no such big fluctuations in the number of passengers traveling within the EU Non-Schengen zone as in the Schengen one.

Figure 5: Historical data- EU Non-Schengen passengers



From the historical data for the ROW passengers, plotted in Figure 6, it also became clear that a good approximation for this passengers group would be to take the average value from the last three years as an annual prediction, because the values for these years are almost equal and no trend was found.

Please note that in month April 2010 there was an eruption of the volcano Eyjafjallajökull in Iceland, which led to untypically low demands in this period. For the prediction the corrected demand of 950,000 (S), 460,000 (NS) and 220,000 (ROW) passengers will be taken.

Figure 6: Historical data- Rest of the World passengers

The planned annual passenger numbers, calculated using the historical data, the corrected demand for the month April 2010 and the predictions for the growth made above, for the three groups are: $x_{P(S)}^{APlan} = 12{,}661{,}589$ for the first group of passengers (Schengen), $x_{P(NS)}^{APlan} = 7{,}198{,}603$ for the second group (Non-Schengen) and $x_{P(ROW)}^{APlan} = 2{,}588{,}332$ for the third group (Rest of the World). How this numbers are calculated is shown in the equations (10) and (11):

- For EU Schengen and EU Non-Schengen passengers:

$$x_{P(q)2012}^{APlan} = x_{P(q)2011}^{AAct} + X\%.\, x_{P(q)2011}^{AAct} = (1 + X\%).\, x_{P(q)2011}^{AAct} \tag{10}$$

- For ROW passengers:

$$x_{P(ROW)2012}^{APlan} = \frac{x_{P(ROW)2011}^{AAct} + x_{P(ROW)2010}^{AAct} + x_{P(ROW)2009}^{AAct}}{3} \tag{11}$$

The index ‘P’ in the bottom stands for passengers, the index ‘q’=S, NS stands for the two groups of passengers- Schengen, Non-Schengen, the ‘APlan’/’AAct’ on the top shows that this is the annual plan/ annual actual value and ‘X’ is the per cent value of the growth that is expected for the next year.

Using the historical data and the formulas stated above ne calculates the values for the single parameters. For every one of the five steps an example will be given to show how the calculation is done. The rest of the numbers as well as the historical data are to be found in the Tables.

1) *Passenger numbers fractions $a_{P(k),M(y)}$ of the different segments, years and months:*

The first step of the calculation is to find out the passenger numbers fractions of the different segments, years and months. To do so one has to take the passenger number, for an exact month and passenger category, and to divide it by the yearly passenger demand for the same category. As an example the value for the month January 2011 of the segment Schengen is calculated in equation (12).

$$a_{P(S),M1(2011)} = \frac{x_{P(S),M1(2011)}}{x_{P(S)(2011)}} = \frac{714{,}211}{12{,}001{,}506} = 5.951\% \tag{12}$$

The passenger demands for the other segments, months and years are to be calculated using the exact same methodology. The results can be found in Table 1.

Table 1: Passenger Numbers fractions $a_{P(k),M(y)}$

	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12
$a_{P(S),M(2011)}$	5.951%	6.060%	7.643%	8.370%	9.394%	9.236%	9.778%	9.538%	9.990%	9.042%	7.647%	7.351%
$a_{P(S),M(2010)}$	5.893%	5.983%	7.604%	8.353%	9.310%	9.252%	9.837%	9.642%	9.847%	9.133%	7.873%	7.275%
$a_{P(S),M(2009)}$	6.207%	6.327%	7.726%	8.495%	9.059%	9.222%	9.783%	9.228%	9.491%	9.087%	7.709%	7.666%
$a_{P(NS),M(2011)}$	5.760%	5.626%	6.990%	7.933%	8.805%	9.165%	10.839%	10.481%	10.048%	9.023%	7.700%	7.629%
$a_{P(NS),M(2010)}$	5.911%	5.804%	7.074%	7.890%	8.649%	9.208%	10.846%	10.669%	9.884%	9.187%	7.730%	7.147%
$a_{P(NS),M(2009)}$	6.363%	6.050%	7.192%	8.050%	8.894%	9.415%	10.531%	10.322%	9.501%	8.948%	7.428%	7.307%
$a_{P(ROW),M(2011)}$	7.381%	6.318%	7.361%	7.985%	7.920%	8.589%	10.671%	9.651%	9.417%	9.240%	7.921%	7.546%
$a_{P(ROW),M(2010)}$	6.968%	6.622%	7.945%	8.268%	8.302%	8.598%	10.409%	10.055%	9.115%	9.198%	7.400%	7.119%
$a_{P(ROW),M(2009)}$	7.093%	6.534%	7.641%	8.581%	7.981%	8.027%	10.289%	10.528%	9.164%	9.176%	7.755%	7.231%

2) *Monthly segment mean values of the passenger numbers $\mu_{a(k),M}$:*

The monthly segment means are calculated by averaging the values for the last years. The monthly segment means of the passenger numbers for segment S and month 1 are calculated explicitly to show the methodology.

$$\mu_{a(S),M1} = \sum_{y=1}^{Y} \frac{a_{P(S),M1(y)}}{Y} = \frac{a_{P(S),M1(2011)} + a_{P(S),M1(2010)} + a_{P(S),M1(2009)}}{3} = \frac{5.951\% + 5.893\% + 6.207\%}{3} = 6.017\% \tag{13}$$

By using the same formula and the values from Table 1 one calculates the following values for the other segments and months.

Table 2: Monthly fraction segment mean values

	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12
$\mu_{a(S),M}$	6.017%	6.123%	7.658%	8.406%	9.254%	9.237%	9.799%	9.469%	9.776%	9.087%	7.743%	7.431%
$\mu_{a(NS),M}$	6.012%	5.827%	7.085%	7.958%	8.782%	9.263%	10.739%	10.491%	9.811%	9.053%	7.620%	7.361%
$\mu_{a(ROW),M}$	7.147%	6.491%	7.649%	8.278%	8.068%	8.405%	10.457%	10.078%	9.232%	9.205%	7.692%	7.299%

After having calculated the monthly segment mean values the next step is to calculate the passenger numbers for each month and group.

3) *Mean values of the fraction-based segment passenger numbers*

As shown in the construction part the mean values of the fraction-based passenger numbers are equal to the product of the annual segment passenger numbers and the monthly segment means of the passenger numbers. Here the example of how the calculation is to be done is again for segment Schengen, year 2011 and month M1:

$$\mu_{P(S),M1} = x_{P(S)}^{Plan} * \mu_{a(S),M1} = 12{,}661{,}589 * 6.017\% = 761{,}840 \quad (14)$$

The values for the other periods and segments can be seen in Table 3 bellow.

Table 3: Mean values of the fraction based segment passenger numbers

	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12
$\mu_{P(S),M}$	761,840	775,284	969,578	1,064,308	1,171,744	1,169,493	1,240,741	1,198,964	1,237,798	1,150,613	980,378	940,847
$\mu_{P(NS),M}$	432,746	419,463	510,034	572,848	632,208	666,807	773,028	755,179	706,247	651,664	548,498	529,882
$\mu_{P(ROW),M}$	184,996	168,009	197,979	214,267	208,816	217,539	270,652	260,857	238,958	238,248	199,099	188,912

4) *The mean value of the fractional single-period portfolio passenger numbers*

The mean value of the fractional single-period portfolio passenger numbers is equal to the sum of the three mean values of the segment passenger numbers. The calculation for fraction one is shown in the next equation:

$$\mu_{P(PF),M1} = \sum_{k} \mu_{P(k),M1} = \mu_{P(S),M1} + \mu_{P(NS),M1} + \mu_{P(ROW),M1}$$
$$= 761{,}840 + 432{,}746 + 184{,}996 = 1{,}379{,}582 \quad (15)$$

Table 4 contains the mean values of the fractional single-period portfolio passenger numbers.

Table 4: Mean values of the fractional single-period portfolio passenger numbers

	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12
$x^{Plan}_{P(PF)}$	1,379,582	1,362,756	1,677,592	1,851,423	2,012,767	2,053,839	2,284,421	2,215,000	2,183,003	2,040,526	1,727,975	1,659,640

5) *The mean value of the multi- period portfolio passenger numbers* $\mu_{P(PF),T}(s_0)$

Here the mean value for the month April is calculated as an example:

$$\mu_{P(PF),T4}(s_0) = \sum_{M \leq T4} \mu_{P(PF),T}$$
$$= \mu_{P(PF),M1} + \mu_{P(PF),M2} + \mu_{P(PF),M3} + \mu_{P(PF),M4} \quad (16)$$

$$= 1{,}379{,}582 + 1{,}362{,}756 + 1{,}677{,}592 + 1{,}851{,}423$$
$$= 6{,}271{,}353$$

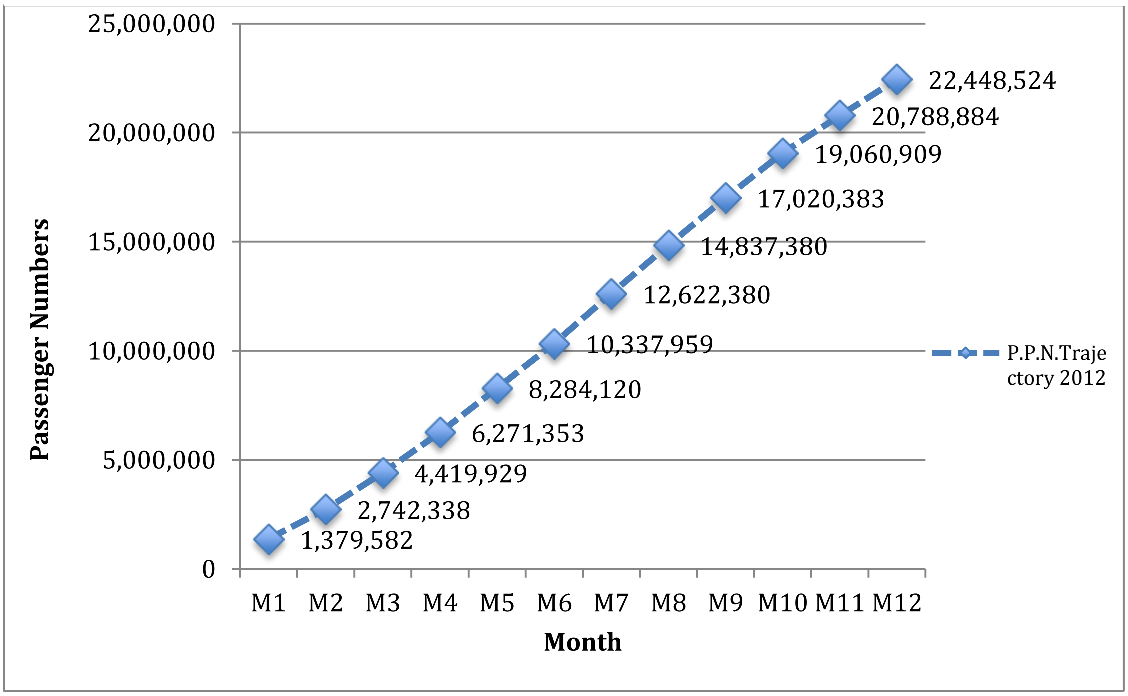
The rest of the values are to be found in the following Table 5:

Table 5: Mean values of the multi-period portfolio passenger numbers

	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12
$\mu_{P(PF),T}(s_0)$	1,379,582	2,742,338	4,419,929	6,271,353	8,284,120	10,337,959	12,622,380	14,837,380	17,020,383	19,060,909	20,788,884	22,448,524

As seen in the last step, the Gaussian model can be used to evaluate the means over multiple sub-periods. The planned passenger numbers trajectory, visualised in Figure 7 by the dashed blue line, is the sequence of all intermediating passenger numbers means from one month to one year. For the first month the number of passengers of the accumulated passenger numbers forecast in the planned passenger numbers trajectory equals the forecast for the month January. For the multi-period time horizons, as shown in the formula above, the monthly passenger number forecasts are summed order to give the intermediate forecast numbers. The annual number of passengers is equal to the sum of all twelve months.

Figure 7: Multi-period planned passenger numbers trajectory



2.3 Multi-period 90% enterprise passenger numbers uncertainty corridor

In this section the uncertainty corridor that accompanies the multi-period planned passenger numbers trajectory computed earlier, will be calculated. This one will consist out of the lower and the upper bounds in which the monthly and annual forecasts will land with a probability of 90%.

2.3.1 Construction of the multi-period passenger numbers volatility

The inclusion of the volatilities allows the integration of uncertainty in the plan activities. In a Gaussian model the standard deviation is the mathematical representation of the enterprise passenger numbers volatility around their forecasted values. As in mathematics the standard deviation is defined as the square root of the variance, which is a measure of the dispersion of a set of data points around their mean, it follows that one can define the volatility as the square root of the variance:

$$\begin{aligned} &StdDev[q] = \sqrt{Var[s]} \\ &or \\ &\sigma = \sqrt{Var[s]} \end{aligned} \tag{17}$$

where σ stands for the volatility and can be interpreted as the average deviation from the mean (*Cleff, 2008 P. 59)*.

In the fraction based Gaussian enterprise model the intermediate enterprise passenger numbers volatilities in form of the s_0- conditional annual passenger numbers volatility parameter $\sigma_{P(PF),T}(s_0)$ is mathematically defined in the following equation (18) as the square root of the summed variances.

$$\sigma_{P(PF),T}(s_0) = \sqrt{\sum_{M \leq T} \sigma^2_{P(PF),M}\,(s_0)} \tag{18}$$

One has to be careful that calculating the needed volatilities is trickier than calculating the mean values. As shown in the formula above, one has to transform these into variances before summing them up! This has to be considered by calculating all variables consisting of volatilities.

In the next equation (19) the calculation of the monthly portfolio volatilities $\sigma_{P(PF),M}$ of the fractional single-period passenger numbers is mathematically defined. In this equation j=i=S, NS, ROW segments and $\rho_{P(i),P(j),M}$ is the correlation between the fraction-based segment passenger numbers $\tilde{x}_{P(k),M}$

$$\sigma_{P(PF),M} = \sum_{i} \sum_{j} \sigma_{P(i),M} \cdot \rho_{P(i),P(j),M} \cdot \sigma_{P(j),M} \tag{19}$$

(Markowitz, 1959 P. 91)

In the above calculation due to the properties of the fraction-based enterprise model the monthly segment passenger numbers volatilities $\sigma_{P(k),M}$ can be defined as the product of the annual segment passenger numbers plan with the monthly segment passenger numbers fraction $\sigma_{a(k),M}$. The mathematical definition is given in the following equation (20):

$$\sigma_{P(k),M} = x_{P(k)}^{Plan} \cdot \sigma_{a(k),M} \tag{20}$$

In equation (19) the correlation can be equal to only numbers between -1 and 1 and can be defined as the coefficient that shows how much the one variable changes when the other does. If the correlation converges to 1 or -1 this is a sign that the variables have a higher positive respectively negative connection to each other. If the correlation coefficient is equal to zero, than one can say that both variables have no linear connection to each other. One can also define the correlation as a function of the covariance as shown in equation (21) *(Cleff, 2008 P. 106-110*)

$$\rho_{a(i),a(j)} = \frac{Cov\left[a_{P(i),M}, a_{P(j),M}\right]}{Vola\left[a_{P(i),M}\right] \cdot Vola\left[a_{P(j),M}\right]} \tag{21}$$

(Cleff, 2008 P. 109)

The covariance is a measure for the deviation of a pair of points from the so called bivariate main point of a scatter diagram, which on the other hand is a measure of how two variables are correlated. Mathematically the covariance is defined as in equation (22):

$$\begin{aligned} &Cov\left[a_{P(i),M}, a_{P(j),M}\right] = \\ &= \frac{1}{Y-1} \sum_{i=1}^{n} \left(a_{P(i),M(y)} - \mu_{a(i),M}\right)\left(a_{P(j),M(y)} - \mu_{a(j),M}\right). \end{aligned} \tag{22}$$

As one can see in the above equation the covariance depends on it`s scale unit. In order to solve this problem as it can be seen in equation (21), where the correlation is mathematically defined, it has to be divided by the standard deviations of the two variables so that one can get a dimensionless coefficient (*Cleff, 2008 P. 106-109)*.

The last variables needed for the calculation of the intermediate enterprise passenger numbers volatilities are the volatilities of the monthly segment passenger numbers fraction $\sigma_{a(k),M}$. These are derived by the naive calibration method as shown in equation (23) and (24):

$$\sigma_{a(k),M} = \max\left(Vola\left[a_{P(k),M}\right]; X\%\right) \quad (23)$$

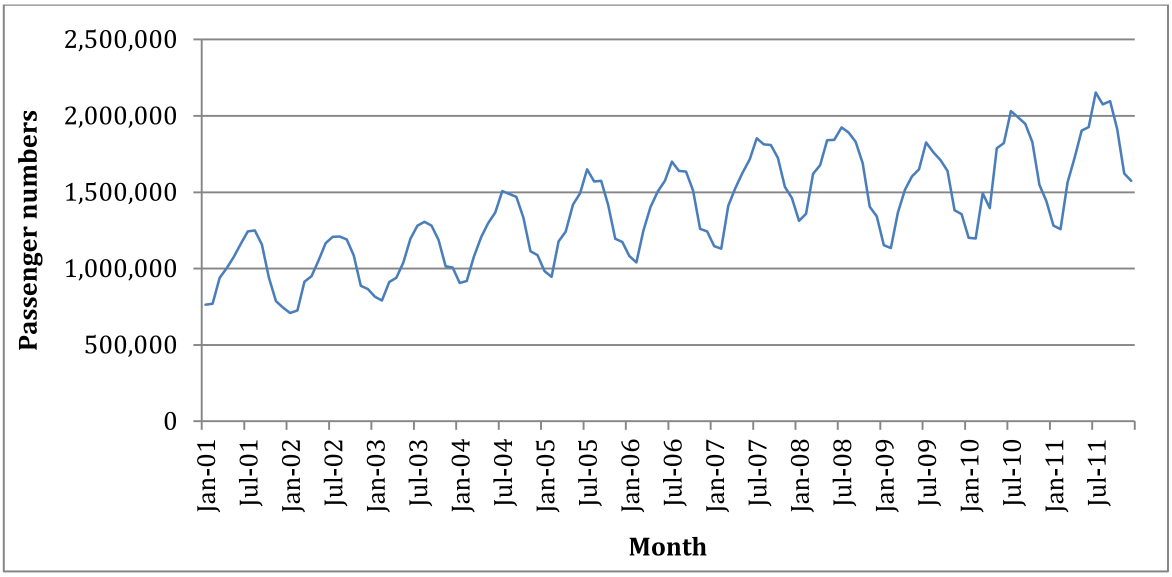
$$Vola\left[a_{P(k),M}\right] = \sqrt{\sum_{y=1}^{Y} \frac{\left(a_{P(k),M(y)} - \mu_{a(k),M}\right)^2}{Y-1}} \quad (24)$$

(Cleff, 2008 P. 60)

As one can see in equation (24) by the calculation of the variance (the part under the square root) one divides not by the number of years but by the number of years minus one. This so-called theoretical variance is used when one has to make a conclusion for the population out of a sample. In such cases only the theoretical variance provides the unbiased estimation of the dispersion from the sample out to the population, when the mean value of the population (expected value) unknown is *(Cleff, 2008 P. 60-61).*

The X% is chosen as a starting lower boundary for the volatility, which might have to be corrected later if the uncertainty corridor is too tight. The lower boundary for the volatility is needed, because every number underneath it wouldn’t describe the reality accurately enough, something that might lead people thinking that this business is very secure and stable, which is not the reality. An example of the data’s fluctuations and uncertainty hidden in this business can be seen at Figure 8, where the historical data for the years between 2001 and 2011 is graphically represented.

Figure 8: Historical data 2001-2011



One can see that after a steady growth of the passenger numbers between 2003 and August 2008, from the month November 2008 the passenger demand became less and in the year 2009 was even below the one for 2007. This accompanied with the case in April 2010 when after a volcano eruption there were no flights for almost two weeks are clear examples of how passenger numbers could fluctuate.

When all the formulas from above are put together this leads to the one for the calculation of the volatilities of the multi-period enterprise passenger numbers:

$$\sigma_{P(PF),T}(s_0) = \sqrt{\sum_{M \leq T} \sum_i \sum_j x_{P(i),M}^{Plan} \cdot \sigma_{a(i),M} \cdot \rho_{a(i),a(j),M} \cdot x_{P(j),M}^{Plan} \cdot \sigma_{a(j),M}} \quad (25)$$

As in the previous section, where the multi-period planned passenger numbers trajectory was defined this section will also continue with the calibration process, but this time of the multi-period passenger numbers volatilities. Again in order to give the reader a better feeling of how the model is used for every step of the calculation a numerical example will be given.

2.3.2 Historical Calibration of the Multi-period planned passenger numbers trajectory

Earlier the passenger numbers fractions of the different segments, years, months $a_{P(k),M(y)}$ and the monthly fraction segment mean values of the passenger numbers $\mu_{a(k),M}$, which can be found in Table (1) and (2) were calculated. By fitting this data into the formulas from above one can calculate the Multi-period passenger numbers volatilities by following the above-mentioned steps. Here again examples will be given to show how the equations and formulas given are to be correctly used.

1) *Calculation of the volatilities of the monthly passenger numbers fractions $\sigma_{a(k),M}$ of the different segments, years and months:*

The volatilities of the monthly passenger numbers fraction are a function of both the passenger number fractions of the different segments and the monthly fraction segment mean values:

$$Vola\left[a_{P(S),M1}\right] = \sqrt{\sum_{y=1}^{Y} \frac{\left(a_{P(S),M1(y)} - \mu_{a(S),M1}\right)^2}{Y-1}} \tag{26}$$

$$= \sqrt{\frac{\left(a_{P(S),M1(2011)} - \mu_{a(S),M1}\right)^2 + \left(a_{P(S),M1(2010)} - \mu_{a(P),M1}\right)^2 + \left(a_{P(S),M1(2009)} - \mu_{a(P),M1}\right)^2}{2}}$$

$$= \sqrt{\frac{(5.951\% - 6.017\%)^2 + (5.893\% - 6.017\%)^2 + (6.207\% - 6.017\%)^2}{2}} = 0.167\%$$

By the same process as used for fragment S month M1 one obtains the following results for the volatilities of the other segments:

Table 6: Monthly passenger numbers fractions $\sigma_{a(k),M}$

	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12
$\sigma_{a(S),M}$	0.167%	0.180%	0.062%	0.078%	0.174%	0.015%	0.033%	0.215%	0.257%	0.045%	0.117%	0.208%
$\sigma_{a(NS),M}$	0.314%	0.213%	0.101%	0.082%	0.124%	0.134%	0.180%	0.174%	0.280%	0.122%	0.166%	0.246%
$\sigma_{a(ROW),M}$	0.212%	0.156%	0.292%	0.298%	0.205%	0.327%	0.195%	0.439%	0.162%	0.032%	0.266%	0.221%

All values that are under 0.1% are corrected to 0,1%, because it is believed that all the values underneath it wouldn't describe the uncertainties accurately enough. As an example the volatility for the Schengen passengers in month 4, $\sigma_{a(S),M4}$ can be given:

$$\sigma_{a(S),M4} = \max\left(Vola\left[a_{s(S),M1}\right]; 0.1\%\right) = \max(0.078\%; 0.1\%) = 0.1\% \tag{27}$$

After doing this for all months and segments one gets the following values, which are also to be used for the further calculation:

Table 7: Corrected monthly passenger numbers fractions $\sigma_{a(k),M}$

	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12
$\sigma_{a(S),M}$	0.167%	0.180%	0.100%	0.100%	0.174%	0.100%	0.100%	0.215%	0.257%	0.100%	0.117%	0.208%
$\sigma_{a(NS),M}$	0.314%	0.213%	0.101%	0.100%	0.124%	0.134%	0.180%	0.174%	0.280%	0.122%	0.166%	0.246%
$\sigma_{a(ROW),M}$	0.212%	0.156%	0.292%	0.298%	0.205%	0.327%	0.195%	0.439%	0.162%	0.100%	0.266%	0.221%

2) *Calculation of the Covariance $Cov[a_{P(i),M}, a_{P(j),M}]$ and correlation $\rho_{a(i),a(j),M}$ between the passenger numbers fractions of the different segments in the different months:*

The covariance and correlation both describe how much the variables differ from the expected value in similar ways. Said in other words these two show how much the one variable changes when the other does. As an example here the covariance between the passenger numbers fraction of segments S and NS in month M1 will be calculated:

$$Cov[a_{P(S),M1}, a_{P(NS),M1}] = \\ = \sum_{y=1}^{Y} \frac{(a_{P(S),M1(y)} - \mu_{a(S),M1}) \cdot (a_{P(NS),M1(y)} - \mu_{a(NS),M1})}{Y-1} = \\ = \frac{\begin{array}{c}(a_{P(S),M1(2011)} - \mu_{a(S),M1}) \cdot (a_{P(NS),M1(2011)} - \mu_{a(NS),M1}) \\ +(a_{P(S),M1(2010)} - \mu_{a(S),M1}) \cdot (a_{P(NS),M1(2010)} - \mu_{a(NS),M1}) \\ +(a_{P(S),M1(2009)} - \mu_{a(S),M1}) \cdot (a_{P(NS),M1(2009)} - \mu_{a(NS),M1})\end{array}}{3-1} = \\ = \frac{\begin{array}{c}(5.951\% - 6.017\%) \cdot (5.760\% - 6.012\%) \\ +(5.893\% - 6.017\%) \cdot (5.911\% - 6.012\%) \\ +(6.207\% - 6.017\%) \cdot (6.363\% - 6.012\%)\end{array}}{2} = 0.00000480 \tag{28}$$

The correlation coefficient between the Schengen and Non-Schengen passenger numbers is a function of the covariance and the volatilities of the both variables and is defined as:

$$\rho_{a(S),a(NS),M1} = \frac{Cov\left[a_{P(S),M1}, a_{P(NS),M1}\right]}{Vola\left[a_{P(S),M1}\right] \cdot Vola\left[a_{P(NS),M1}\right]} = \frac{0.00000480}{0.0167\% \cdot 0.314\%} = 0.914 \tag{29}$$

The other values for the correlations between the different segments in the different months are:

Table 8: Correlations $\rho_{a(i),a(j),M}$ between the passenger numbers fractions

	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12
$\rho_{a(S),a(NS),M}$	0.91393	0.79797	0.73643	0.98863	-0.60021	-0.77048	0.46099	0.94749	0.99988	0.67102	0.33620	-0.00793
$\rho_{a(S),a(ROW),M}$	-0.05005	0.02276	-0.33484	0.82196	0.12727	0.86920	-0.27305	-0.74977	0.60893	-0.64312	-0.99899	-0.08101
$\rho_{a(NS),a(ROW),M}$	-0.45110	0.62070	0.39087	0.72698	-0.86973	-0.98490	0.72781	-0.49879	0.62113	0.13622	-0.29351	0.99732

3) *Calculation of the planned volatilities of the fraction single-period enterprise passenger numbers $\sigma_{P(PF),M}$:*

As an example here the values for month M1 will be calculated explicitly:

$$\sigma_{P(PF),M1} = \sqrt{\sum_i \sum_j \sigma_{P(i),M1} \cdot \rho_{P(i),P(j),M1} \cdot \sigma_{P(j),M1}} \tag{30}$$

$$= \sqrt{\begin{array}{c} \sigma^2_{P(S),M1} + \sigma^2_{P(NS),M1} + \sigma^2_{P(ROW),M1} + 2 \cdot \sigma_{P(S),M1} \cdot \rho_{P(S),P(NS),M1} \cdot \sigma_{P(NS),M1} \\ +2 \cdot \sigma_{P(S),M1} \cdot \rho_{P(S),P(ROW),M1} \cdot \sigma_{P(ROW),M1} + 2 \cdot \sigma_{P(NS),M1} \cdot \rho_{P(NS),P(ROW),M1} \cdot \sigma_{P(ROW),M1} \end{array}}$$

$$= \sqrt{\begin{array}{c} \left(x^{Plan}_{P(S)} \cdot \sigma_{a(S),M1}\right)^2 + \left(x^{Plan}_{P(NS)} \cdot \sigma_{a(NS),M1}\right)^2 + \left(x^{Plan}_{P(ROW)} \cdot \sigma_{a(ROW),M1}\right)^2 \\ +2 \cdot x^{Plan}_{P(S)} \cdot \sigma_{a(S),M1} \cdot \rho_{a(S),a(NS),M1} \cdot x^{Plan}_{P(NS)} \cdot \sigma_{a(NS),M1} \\ +2 \cdot x^{Plan}_{P(S)} \cdot \sigma_{a(S),M1} \cdot \rho_{a(S),a(ROW),M1} \cdot x^{Plan}_{P(ROW)} \cdot \sigma_{a(ROW),M1} \\ +2 \cdot x^{Plan}_{P(NS)} \cdot \sigma_{a(NS),M1} \cdot \rho_{a(NS),a(ROW),M1} \cdot x^{Plan}_{P(ROW)} \cdot \sigma_{a(ROW),M1} \end{array}}$$

$$= \sqrt{\begin{array}{c} (12{,}661{,}589 \cdot 0.167\%)^2 + (7{,}198{,}603 \cdot 0.314\%)^2 + (2{,}588{,}332 \cdot 0.212\%)^2 \\ +2 \cdot 12{,}661{,}589 \cdot 0.167\% \cdot 0.91393 \cdot 7{,}198{,}603 \cdot 0.314\% \\ +2 \cdot 12{,}661{,}589 \cdot 0.167\% \cdot (-0.05005) \cdot 2{,}588{,}332 \cdot 0.212\% \\ +2 \cdot 7{,}198{,}603 \cdot 0.314\% \cdot (-0.451098) \cdot 2{,}588{,}332 \cdot 0.212\% \end{array}}$$

$$= 41{,}722$$

By making the same calculation for the other months one gets the following values, shown in Table 9, for the rest of the volatilities:

Table 9: Volatilities of the fraction single-period enterprise passenger numbers $\sigma_{P(PF),M}$

	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12
$\sigma_{P(PF),M}$	41,722	37,611	19,641	26,325	17,499	12,764	23,767	32,702	55,373	18,933	16,651	34,742

4) *Calculation of the volatilities of the multi-period enterprise passenger numbers $\sigma_{P(PF),T}$:*

On the next page an example is given of how the volatility of the multi-period enterprise passenger numbers for the month April is calculated:

$$\sigma_{P(PF),T4}(s_0) = \sqrt{\sum_{Q \leq T} \sigma^2_{P(PF),M}}$$
$$= \sqrt{\sigma^2_{P(PF),M1} + \sigma^2_{P(PF),M2} + \sigma^2_{P(PF),M3} + \sigma^2_{P(PF),M4}} \tag{31}$$
$$= \sqrt{41{,}722^2 + 37{,}611^2 + 19{,}641^2 + 26{,}325^2} = 65{,}070$$

By calculating the values for the other months in the same manner one gets:

Table 10: Volatilities of the multi-period enterprise passenger numbers $\sigma_{P(PF),T}$

	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12
$\sigma_{P(PF),T(s_0)}$	41,722	56,172	59,507	65,070	67,381	68,580	72,581	79,608	96,972	98,803	100,197	106,049

2.3.3 Construction of the multi-period enterprise passenger numbers VaRs

As one of the references states: “Value-at-Risk (VaR) measures the worst expected loss under normal market conditions over a specific time interval at a given confidence level” *(Benninga S. and Wiener Z, 1998 P. 1)*. The VaR can also be seen as the lowest quantile of the possible losses that can ensue within a specified portfolio during a given time period.

If one wants to answer the question what is the value X under that the passenger numbers are not going be, with the possibility of Y%, than one is asking for the Y quantile of the normal distribution $N(\mu;\sigma^2)$. Mathematically this is defined the following way: If $X \sim N(\mu;\sigma^2)$ and $X^* \sim N(0;1)$ than the α- Quantile τ_α respectively τ_α^* are defined the following way:

$$E[X \leq \tau_\alpha] = P[X^* \leq \tau_\alpha^*] = \alpha \qquad (32)$$

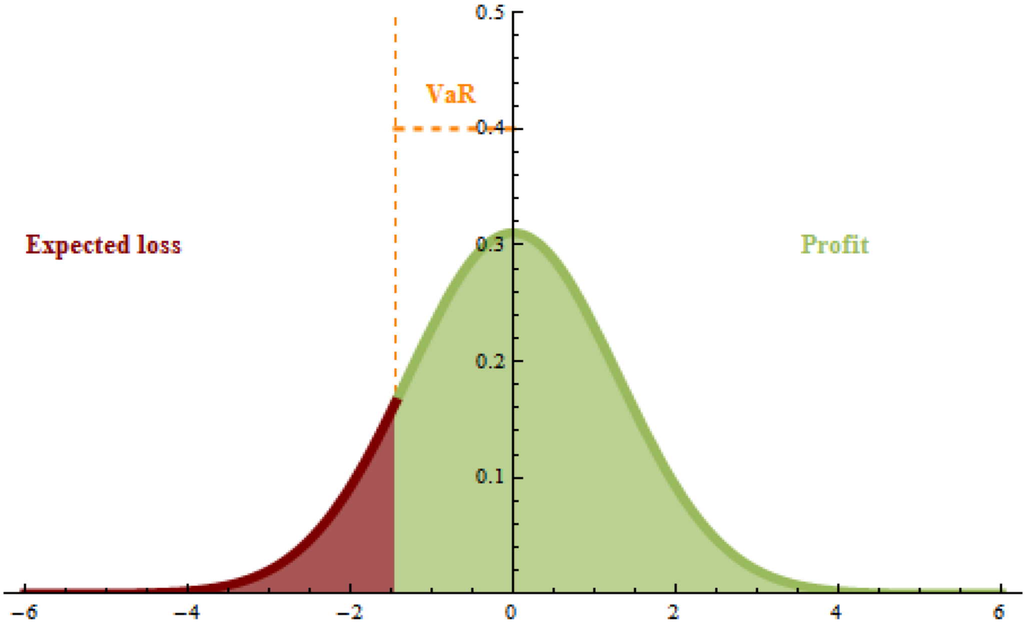
And

$$\tau_\alpha = \mu + \sigma\tau_\alpha^* \qquad (33)$$

(Arens, Hettlich, Karpfinger, Kockelhorn, Lichtenberg and Stachel, 2008 P. 1350).

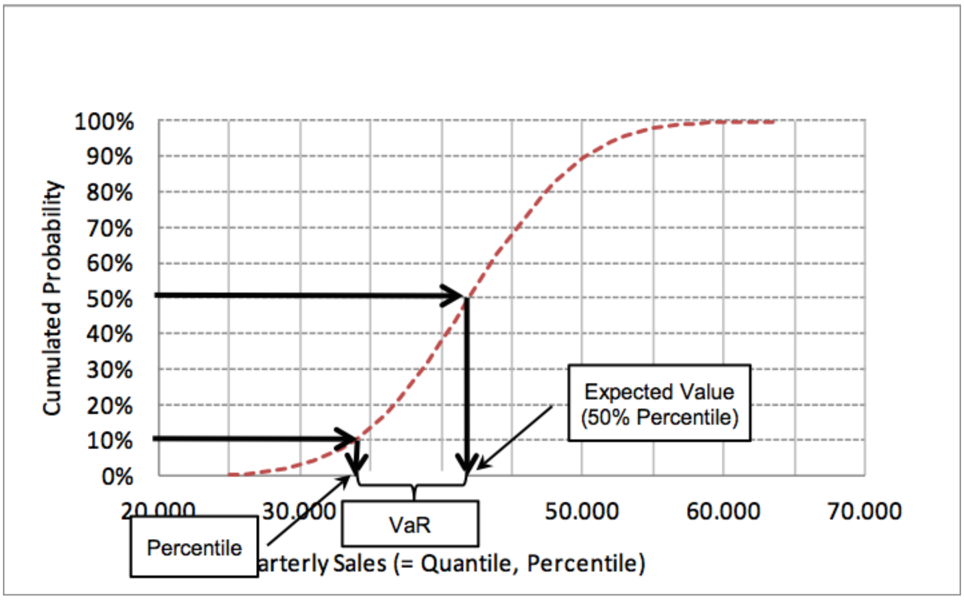
The graphical representation of the VaR can be seen in the next two figures.

Figure 9: VaR, Normal Density Function



Source: http://demonstrations.wolfram.com/ValueAtRisk/

Figure 10: VaR, Normal Distribution Function



Source: *Schwaiger, Bös and Kronfellner, 2012 P. 67*

In the case of a standard normally distributed random variable X^* is the probability that a realisation will be under τ_α^* exactly α. As the density of the $N(0;1)$is symmetrical the probability that the realisation is higher than $\tau_{1-\alpha}^*$ also α. On the left hand side of τ_α^* lies the probability α and on the ride hand side the probability $1-\alpha$. The probability between τ_α^* and $\tau_{1-\alpha}^*$ is therefore equal to $1-2\alpha$ *(Arens, Hettlich, Karpfinger, Kockelhorn, Lichtenberg and Stachel, 2008 P. 1350).*

In the second figure the VaR is described in the Normal Distribution Function. Here one can see the probabilities for every value and percentile (please note that the value written on the x-axis are not the real passenger numbers values that were calculated earlier).

Out of the initial passenger numbers volatility the passenger numbers value at risk, needed for the calculation of the confidence corridors, are calculated. For this purpose the annual passenger numbers volatilities are multiplied with the percentiles of the standard normal distribution, those values can be found in Table 11, for the α- probability and the $(1-\alpha)$ probability. The mathematical expression is shown in equation (34)

$$VaR_{P(PF),T}^{1-\alpha}(s_0) = \sigma_{P(PF),T}(s_0) \cdot z^{1-\alpha} \tag{34}$$

2.3.4 Historical calibration of the multi-period enterprise passenger numbers VaRs

The confidence level gives how high is the probability that the values of $\tilde{x}_{P(PF),T}(s_0)$ are between $\mu_{P(PF),T}(s_0)$ and $\mu_{P(PF),T}(s_0) + VaR^{1-\alpha}_{P(PF),T}(s_0)$. In this paper the Values at Risk are calculated for the confidence level of 95%.

Specifically calculated for 95% the VaRs for T3 are:

$$VaR^{95\%}_{P(PF),T3}(s_0) = \sigma_{P(PF),T3}(s_0) \cdot z^{95\%} = 59{,}507 \cdot 1.6449 = 97{,}883 \qquad (35)$$

The other values calculated using the same formula are:

Table 11: Multi-period enterprise passenger numbers VaRs

	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12
$VaR^{95\%}_{P(PF),T}(s_0)$	68,628	92,397	97,883	107,033	110,836	112,807	119,389	130,948	159,510	162,522	164,814	174,440
$VaR^{5\%}_{P(PF),T}(s_0)$	-68,628	-92,397	-97,883	-107,033	-110,836	-112,807	-119,389	-130,948	-159,510	-162,522	-164,814	-174,440

2.3.5 Multi-period 90% enterprise passenger numbers– uncertainty corridor

The calculation of the upper- and lower limits of the multi-period 90% passenger numbers-uncertainty corridor, which boundaries are the initial 5%-percentile and the initial 95%-percentile of the normal distribution $(95\% - 5\% = 90\%)$, is shown in equations (36) and (37).

$$x^{95\%}_{E(PF),T}(s_0) = \mu_{E(PF),T}(s_0) + VaR^{95\%}_{E(PF),T}(s_0) \qquad (36)$$

$$\begin{aligned} x^{5\%}_{E(PF),T}(s_0) &= \mu_{E(PF),T}(s_0) - VaR^{95\%}_{E(PF),T}(s_0) \\ &= \mu_{E(PF),T}(s_0) + VaR^{5\%}_{E(PF),T}(s_0) \end{aligned} \qquad (37)$$

The values of $z^{5\%}$ und $z^{95\%}$ just differ by their algebraic sign. All z values, which are under 50%, have a minus algebraic sign and all z values, which are over 50%, are positive. The z value of 50% equals zero. From the table below one gets $z^{95\%} = 1.6449$ and $z^{5\%} = -1.6449$. In Table 18 shown in the Appendix all the z-values for the different uncertainties are given.

For T3 the corridor is to be calculated as it follows:

$$\begin{aligned} x^{95\%}_{P(PF),T3}(s_0) &= \mu_{P(PF),T3}(s_0) + VaR^{95\%}_{P(PF),T3}(s_0) \\ &= 4{,}419{,}929 + 97{,}883 = 4{,}517{,}812 \end{aligned} \qquad (38)$$

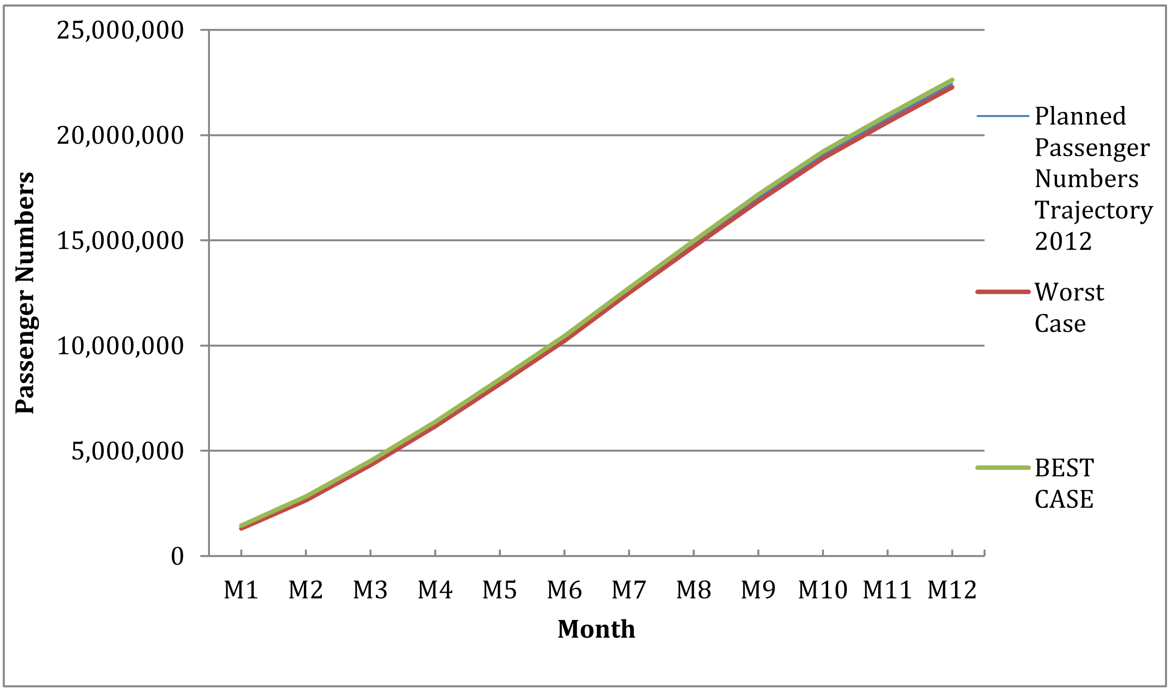
$$x^{5\%}_{P(PF),T3}(s_0) = \mu_{P(PF),T3}(s_0) + VaR^{5\%}_{P(PF),T3}(s_0) = 4{,}419{,}929 - 97{,}883 = 4{,}322{,}047 \quad (39)$$

With a probability of 95% is the value of $\tilde{x}_{P(PF),T3}(s_0)$ fewer than 4,517,812 and with a probability of 5% it will be under 4,322,047. The multi-period planned enterprise passenger numbers trajectory therefore lays between the two values with a probability of 90%. For the other values the following numbers are calculated:

Table 12: Multi-period 90% enterprise passenger numbers uncertainty corridor

	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12
5%	1.448.210	2.834.735	4.517.812	6.378.386	8.394.956	10.450.766	12.741.769	14.968.328	17.179.893	19.223.430	20.953.697	22.622.964
95%	1.310.954	2.649.941	4.322.047	6.164.320	8.173.284	10.225.152	12.502.991	14.706.432	16.860.872	18.898.387	20.624.070	22.274.084

Figure 11: Multi-period 90% enterprise passenger numbers uncertainty corridor



As it can be seen Figure 11 the best and worst case differ very little from the planned passenger numbers trajectory, which means that the uncertainty corridor is too tight. This tight uncertainty corridor, such as a too low volatility, might lead to the false feeling that the air passenger transport is a very secure business, which was shown not to be the reality. The tightness of the corridor shows that the volatilities have chosen are just too low and therefore they will be raised to 1%. The 1% is chosen as an example. If one wants a more conservative plan than one can chose 2%, 3% or even higher depending on how big the fluctuations on the current airport are.

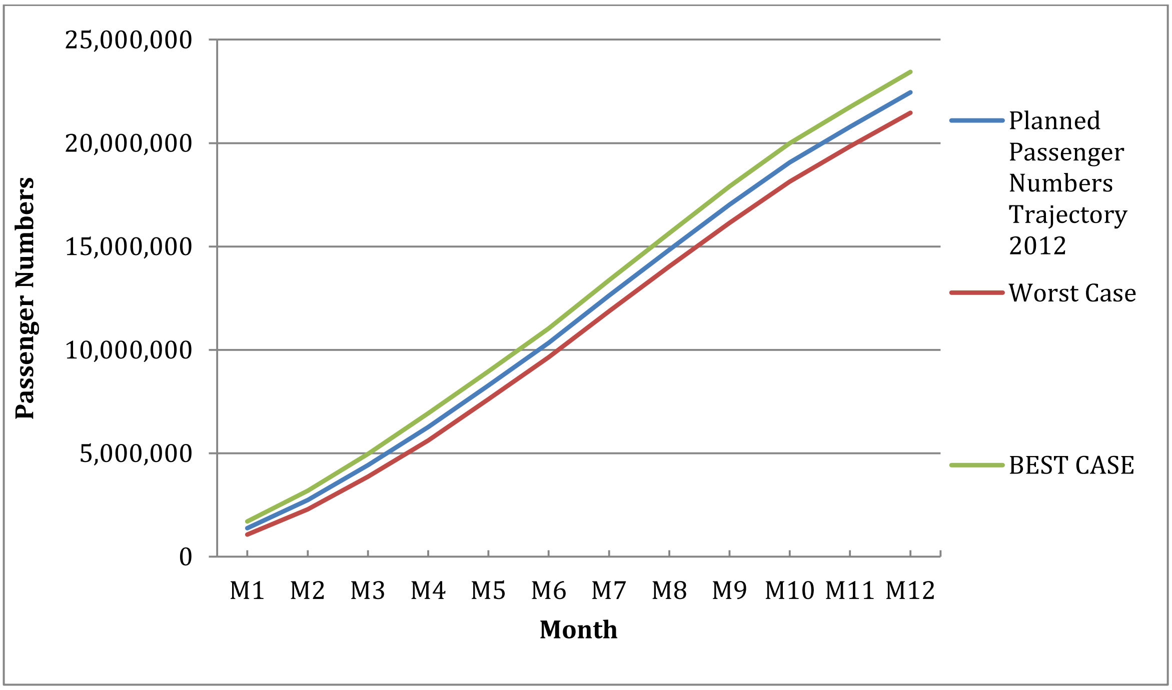
By using the same formulas and α-percentiles, which were used above and the volatility of 1% one gets the following numbers for the multi-period 90% enterprise passenger numbers uncertainty corridor:

Table 13: Multi period 90% enterprise passenger numbers uncertainty corridor for Vola min= 1%

	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12
5%	1.693.989	3.194.092	4.965.473	6.925.185	8.955.380	11.028.815	13.371.864	15.643.198	17.900.684	19.986.763	21.744.039	23.438.352
95%	1.065.175	2.290.584	3.874.386	5.617.521	7.612.859	9.647.102	11.872.896	14.031.562	16.140.081	18.135.055	19.833.728	21.458.696

As it can be seen in Figure 12 our assumption, that the volatility of 0,1% is too low is true and the minimum for the volatility should be set to at least 1%. This is so because a too tight corridor with low volatilities means that if an event happens such as the financial crisis in 2009 then one wouldn't be prepared for such low incomes and this is something one wants to avoid when making forecasts for future events.

Figure 12: Multi-period 90% enterprise passenger numbers uncertainty corridor-new



After having shown an excellent tool for illustrating the possible risks and the passenger numbers objectives has been set, at the next part the model will be validated in an out of sample test where the objectives set will be compared to the real realisations for the year 2012.

2.4 Model validation with an out of sample test of the initial forecasted passenger numbers and uncertainty corridor

In this last part of the second chapter of this work the actual passenger of year 2012 will be compared to the forecast made using the Gaussian Enterprise Model. This out of sample test will show us if the forecast and the uncertainty corridor calculated are accurate and if this model is appropriate for future use.
The historical monthly numbers for year 2012 are:

Table 14: Historical data for year 2012

	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12
$x_{P(PF)}$	1,397,692	1,387,505	1,700,590	1,889,866	1,985,121	2,062,155	2,192,650	2,138,588	2,173,65	1,990,243	1,658,971	1,588,696

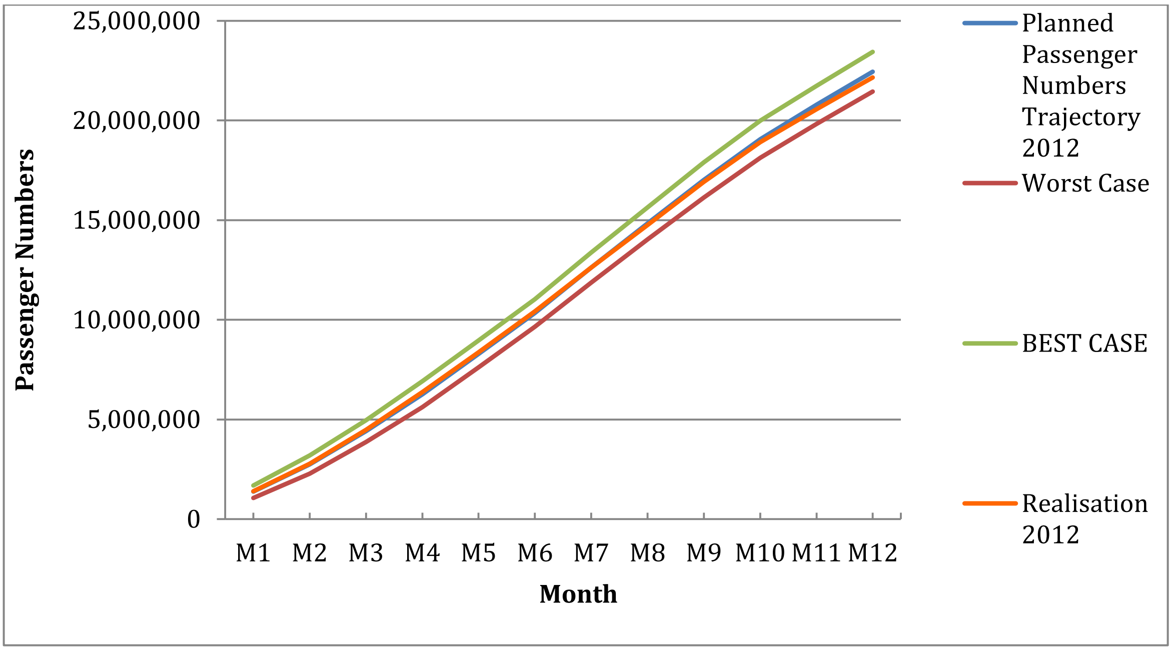
And the accumulated are:

Table 15: Historical data of the multi-period portfolio passenger numbers

	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12
$\mu_{P(PF),T}(S_0)$	1,379,582	2,785,197	4,485,787	6,375,653	8,360,774	10,422,929	12,615,579	14,754,167	16,927,823	18,918,066	20,577,037	22,165,733

In Figure 13 on the next page one can see that the forecast fits very appropriately the historical data and that the uncertainty corridor for a minimum volatility of 1% is also very accurate. The fact that the blue line, which shows our planned path, is almost identical with the orange one, which shows the historical data numbers is a clear indicator of the good quality of the Gaussian Enterprise Model. However, one of the most important aims an enterprise should have is to fit its predictions within the confidence intervals. Being able to do so would mean that the company has effective risk management and place against external factors and unexpected events. This would also mean that the firm not only would not be vulnerable but also would take advantage of stress situations.

Figure 13: Out of sample test for the multi-period portfolio passenger numbers



The Gaussian Passenger Traffic Model is not only used in the initial planning but also in the subsequent planning process in the different states s_t at all time points t. This one among with suggestions for control activities, when the plan/forecast deviations are too high will be shown in the next final part of this section.

2.5 Subsequent Planning

As one of the main ideas of forecasting is the minimization of risks, after a plan is set up one has to check over defined time periods (monthly wise) how the passenger numbers really develop. The s_t- conditional subsequent passenger numbers forecast for the annual portfolio passenger numbers volumes is shown in equation (40)

$$\begin{aligned} \mu_{P(PF),T}(s_t) &= \sum_{M \leq T} x_{P(PF),M} + \sum_{M>T} \mu_{P(PF),M} \\ &= \sum_{M \leq T} x_{P(PF),M} + x_{P(PF)}^{Plan} \cdot \mu_{a(PF),T}(s_t) \end{aligned} \tag{40}$$

Here the conditional rolling forecast $\mu_{P(PF),T}(s_t)$ is calculated out of two components. The first one contains the accumulated actuals resulting out of the previous and current periods $M \leq T$. The second one describes the conditional remaining forecast and is equal to the product of the annual portfolio passenger numbers volume $x_{P(PF)}^{Plan}$ and the sum of the remaining monthly passenger numbers fraction means $\mu_{a(PF),M}(s_t)$.

Table 19, which due to its size is attached in the Appendix, contains of two tables and shows the calculation of the conditional passenger numbers forecasts for the month January 2012 and February 2012 respectively. These tables contain the monthly passenger numbers realizations for the year 2012 and the monthly passenger numbers fraction means, which as mentioned are used in the fraction-based passenger model to compute the subsequent annual passenger numbers forecast. The calculation of the s_1- conditional subsequent passenger numbers forecast, see Table 19, is done by the above stated methodology and is shown in the following equation (41):

$$\begin{aligned} E[\tilde{x}_{P(PF)} | s_1] &= \sum_{M \leq 1} x_{P(PF),M} + \sum_{M>1} \mu_{P(PF),M} \\ &= x_{P(PF),M1} + x_{P(PF)}^{Plan} \cdot \left(\mu_{a,M2}(s_1) + \mu_{a,M3}(s_1) + \cdots + \mu_{a,M12}(s_1)\right) \\ &= 1.397.692 + 22.448.524 \cdot (6{,}07\% + 7{,}47\% + \cdots + 7{,}39\%) \\ &= 1.397.692 + 22.448.524 \cdot 93.85\% \\ &= 1.397.692 + 21.068.942 \\ &= 22.466.634 \end{aligned} \tag{41}$$

Please note that in the above calculation the s_1- conditional subsequent annual passenger numbers forecast, the values for the s_1- conditional fractions means are equal to the s_0-conditional passenger numbers fractions means, which were calculated earlier in 2.2 This is due to the assumption that the conditional passenger numbers fractions are constant over time.

In Table 19 the planned single-period portfolio monthly fraction volatilities are calculated as well. These are functions of the planned monthly segment fraction volatilities, the correlations and the monthly weights of the segments. As one can see in equation (44) the calculation of the planned single-period portfolio monthly fraction volatilities is very similar to the calculation of the planned volatilities of the fraction single-period enterprise passenger numbers $\sigma_{P(PF),M}$. The only difference lies within the fact that one doesn't take the planned fraction passenger segment numbers for the computation but the planned fraction passenger weights $u_{P(k)}^{Plan}$, which are assumed to be constant over time and are equal to the planned annual segment passenger numbers divided by the planned enterprise annual passenger numbers, a definition shown in equation (42).

$$u_{P(k)}^{Plan} = \frac{x_{P(k)}^{Plan}}{x_{P(PF)}^{Plan}} \tag{42}$$

The weights of the Schengen passengers are therefore equal to 56,403%:

$$u_{P(S)}^{Plan} = \frac{x_{P(S)}^{Plan}}{x_{P(PF)}^{Plan}} = \frac{12.661.589}{22.448.524} = 56{,}403\% \tag{43}$$

$$\sigma_{a(PF),M} = \sqrt{\sum_i \sum_j \sigma_{a(i),M} \cdot \rho_{P(i),P(j),M} \cdot \sigma_{a(j),M}} \tag{44}$$

$$= \sqrt{\begin{array}{c} \left(u_{P(S)}^{Plan} \cdot \sigma_{a(S),M}\right)^2 + \left(u_{P(NS)}^{Plan} \cdot \sigma_{a(NS),M}\right)^2 + \left(u_{P(ROW)}^{Plan} \cdot \sigma_{a(ROW),M}\right)^2 \\ +2 \cdot u_{P(S)}^{Plan} \cdot \sigma_{a(S),M} \cdot \rho_{a(S),a(NS),M} \cdot u_{P(NS)}^{Plan} \cdot \sigma_{a(NS),M} \\ +2 \cdot u_{P(S)}^{Plan} \cdot \sigma_{a(S),M} \cdot \rho_{a(S),a(ROW),M} \cdot u_{P(ROW)}^{Plan} \cdot \sigma_{a(ROW),M} \\ +2 \cdot u_{P(NS)}^{Plan} \cdot \sigma_{a(NS),M} \cdot \rho_{a(NS),a(ROW),M} \cdot u_{P(ROW)}^{Plan} \cdot \sigma_{a(ROW),M} \end{array}}$$

As by the calculation of the volatilities of the multi-period enterprise passenger numbers $\sigma_{P(PF),T}$ the s_t- conditional annual enterprise volatility is equal to the square root out of the sum of the s_t- conditional planned monthly fraction variances:

$$\sigma_{a(PF),T}(s_t) = \sqrt{\sum_{M>t} \sigma^2_{a(PF),M}(s_t)} \tag{45}$$

In the above stated formula one calculates the s_1- conditional annual enterprise volatility:

$$\sigma_{a(PF),T}(s_t) = \sqrt{\sum_{M>t} \sigma^2_{a(PF),M}(s_t)} \\ = \sqrt{(0{,}879\%)^2 + (0{,}828\%)^2 + \cdots + (0{,}703\%)^2} = 2{,}54 \tag{46}$$

For the calculation of the s_t- conditional annual passenger numbers volatility one has to multiply the planned annual passenger numbers with the s_t- conditional annual enterprise volatility.

$$\sigma_{P(PF),T}(s_t) = x^{Plan}_{P(PF)} \cdot \sigma_{a(PF),T}(s_t) \tag{47}$$

For the s_t- conditional annual passenger numbers volatility one gets:

$$\sigma_{P(PF),T}(s_t) = x^{Plan}_{P(PF)} \cdot \sigma_{a(PF),T}(s_t) = 22.448.524 * 2{,}39\% = 535.428 \tag{48}$$

Although the aim of this work is to plan the passenger numbers at the Vienna International Airport, here a brief explanation of how a plan/forecast- comparison can be used to specify the probabilities that the planned annual passenger numbers will be realized over the year will be given.

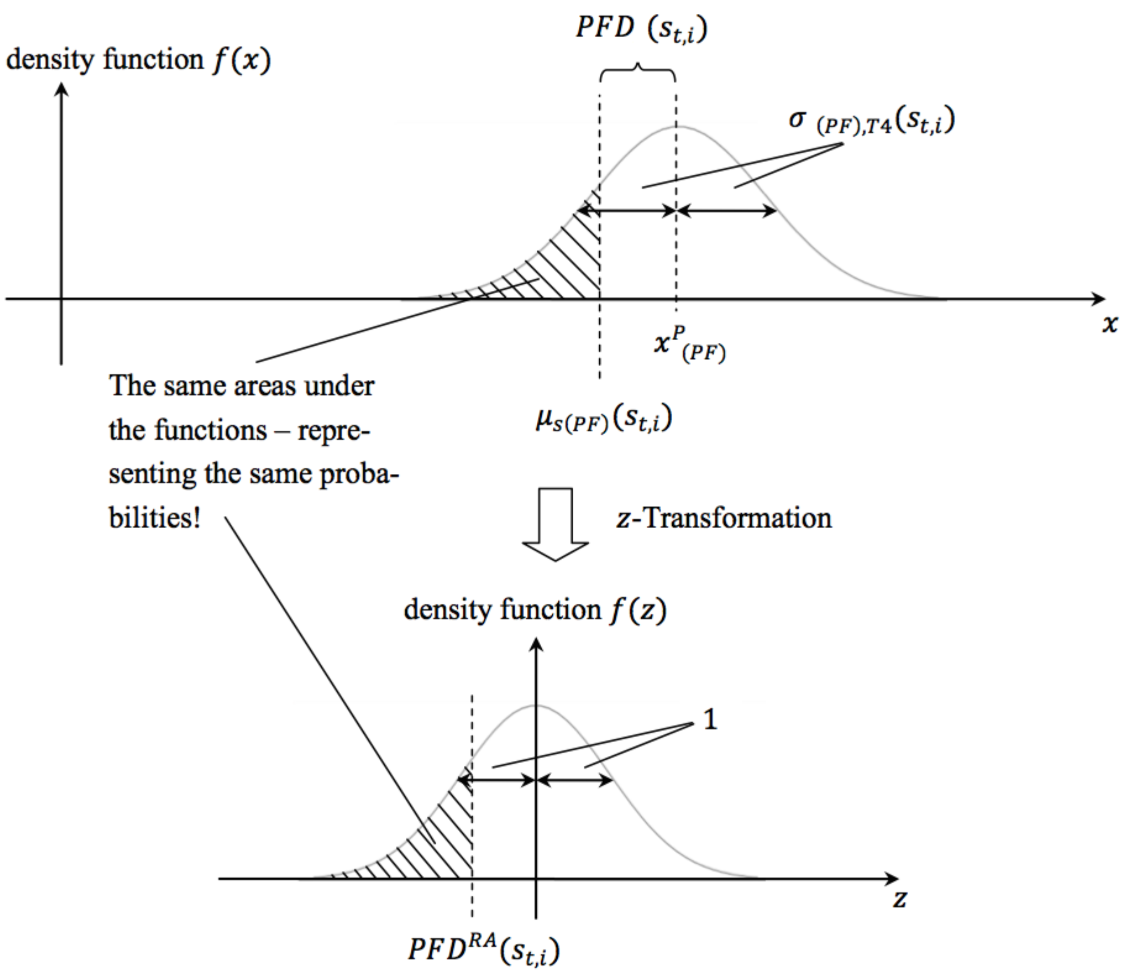
As one can see in Table 19, next to the conditional passenger numbers forecasts and the conditional remaining passenger numbers volatilities also the plan/forecast deviations are computed by subtracting the planned annual enterprise passenger numbers from the conditional forecast, calculation shown in equation (49)

$$PFD_{P(PF)}(s_t) = \mu_{P(PF)}(s_t) - x^{Plan}_{P(PF)} \tag{49}$$

By dividing the PFD by the conditional remaining volatility one gets the Risk Adjusted PFD, which is also called as the z-transformation and is shown in Figure 14 on the next page. This is the transformation of a normal distribution in a resulting standard normal distribution (expected value of zero and the volatility of one), where the probabilities of each value are known.

$$z = \frac{x-\mu}{\sigma} = \frac{\mu_{P(PF)}(s_t) - x_{P(PF)}^{Plan}}{\sigma_{P(PF),T}(s_t)} = \frac{PFD_{P(PF)}(s_t)}{\sigma_{P(PF),T}(s_t)} = PFD^{RA}(s_t) \qquad (50)$$

Figure 14: The z-Transformation



As there is no deviation at the beginning, when the initial planned passenger numbers are calculated, the z-value is zero and the probability is therefore 50%. After month 1 the first actual passenger numbers are realised. The realisation of 1.397.692 travellers is with 18.110 passengers above the planned passenger numbers and therefore the z-value is positive and the p-value, the probability of achieving the annual plan, is now higher than 50%. This so-called check activity after the first month leads to no necessary actions. By taking a closer look at the realizations for month 2 one will see that they are again above the planned values and therefore the expectation of achieving the set goals gets higher again and as by the first month no action is needed. For p-value bellow 50% Schwaiger, Bös and Kronfellner are suggesting the following activities:

<50%: Observation
<30%: Promotion
<15%: Emergency Plan.

The Gaussian Enterprise Model is a very useful tool for one to obtain a better understanding of the numbers and possible outcomes in the forthcoming year. It allows the user not only to make a fraction-based forecast and to include the uncertainties surrounding the passenger numbers volume, resulting in the computation of ‘best and worst’ scenario but also to compare the passenger numbers realisations with the planned passenger numbers and to take quick actions if needed.
After analysing the data using the Gaussian Enterprise Performance Management, the next part of this work will provide a forecast through the Box-Jenkins ARIMA methodology, which is a “typical way” of forecasting airport passenger demands.

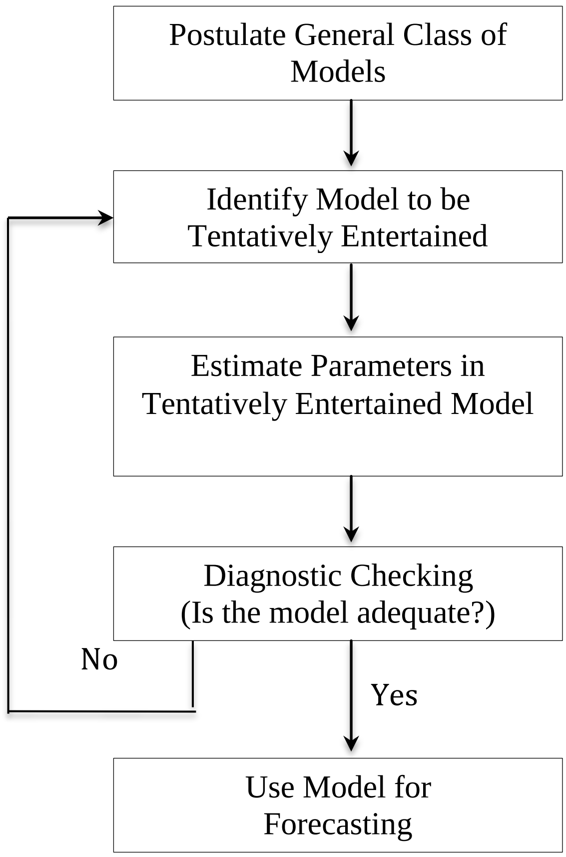
3. The Box-Jenkins/ ARIMA Methodology

There is a variety of statistical forecasting approaches that differ mainly by the data that is used, thus it is crucial to choose the best model and approach when dealing with such tasks. One of the most common approaches includes exponential smoothing, correlation and regression analysis and time series and decomposition analysis. The common factor in the above methods is the assumptions of the values of the series, which are statistically independent or not related to each other. Another popular class of models are those that can produce forecasts based only on a synthesis of historical patterns in data. A specialised subclass includes the Autoregressive integrated moving-average models (ARIMA). Those are linear filtering techniques that completely ignore the independent variables in a forecasting model. In other words, ARIMA models rely completely on the historic and present data of the dependent variable while ignoring the potential external factors – the descriptive variables. It is a highly refined curve-fitting device, which produces the final result of an accurate forecast. In comparison to the first set of models above, ARIMA methodology is adopted when the observations of a time series are statistically dependent on or related to each other. ARIMA forecasts are widely used for forecasts across many different industries. This work will now continue with an explanation of the Box- Jenkins ARIMA Methodology.

3.1 The Box-Jenkins technique

The main feature of the Box-Jenkins forecasting model is the fact that it does not assume any particular pattern in the historic data that is to be analysed and forecasted. Based on a number of factors, it adapts an iterative approach in order to identify the best suiting forecasting model for the set of data. The best possible model is then checked against the historic data in order to analyse to what degree the model accurately describes the series. The appropriateness of the model is assessed based on the residuals between the forecasting model and the historic data. A residual is a measure of the deviation of an observed value in relation to the "theoretical value". In other words the residual is the difference between the observed value and the estimated function value. A model is considered to be appropriate if the residuals between the forecasts and the historic data are small, randomly distributed and independent. Those features of the residual enforce the characteristic of the model – no particular pattern. If the analysis of the residuals shows that the model does not fit and it is not satisfactory, the process is repeated until a more appropriate model is developed. Those steps are also clearly illustrated in the figure below (*Hanke and Reitsch, 1998, P. 407-408*).

Figure 14: Flow diagram of the Box-Jenkins method



Source: Hanke, J. and Reitsch, A., 1998, Business Forecasting P. 408

As shown in the flow chart the analyst can follow those steps of the Box-Jenkins method and choose the most appropriate model for the set of data. A general class of such models include AR, MA and ARIMA. ARs are models with only autoregressive terms; in other words, the output variable depends linearly on its own previous values. MAs are models with only moving-average terms and it is more complicated than AR because the lagged error terms are also observable. The main difference between them is the inclusion of lagged terms as AR includes lagged terms of the series itself and MA includes lagged terms on the residual. The ARIMA model includes both the autoregressive and moving average terms. All of the above are used for stationary time series data. This is a set of data whose average value is not changing over time (*Hanke and Reitsch, 1998, P. 408*).
The Box-Jenkins methodology allows the analyst to select one of the models that best fits the data. The selection of the most appropriate model is done by comparing the distributions of autocorrelation coefficients of the series while fitted with the theoretical distribution for the various models under investigation. The following figures below represent the theoretical distributions for the autocorrelation coefficients for some of the most popular ARIMA models. The autocorrelation function is a set of correlation coefficients between the series and its lagged values over time. The distributions below are highly theoretical and an actual set of data would not produce such a clear and distinctive one. However an experienced analyst should be able to match the produced data to one of the models and identify the best approach. An analysis of the autocorrelation and the partial autocorrelation distributions will be provided further in the work in the discussion of each model (*Hanke and Reitsch, 1998, P. 408).*

Autoregressive model (AR)

The autoregressive model takes the below form:

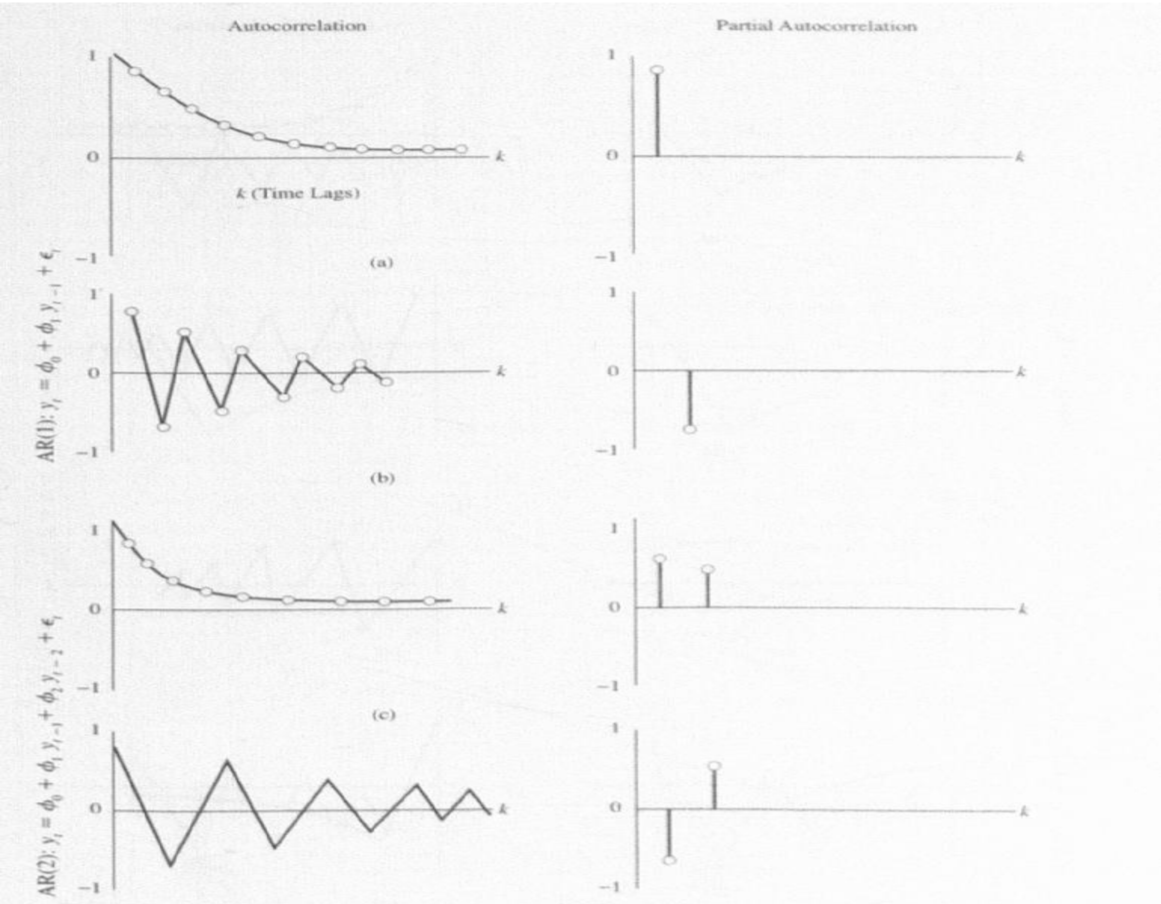
$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \epsilon_t \tag{51}$$

where

- Y_t = Dependent variable
- $Y_{t-1}, Y_{t-2}, Y_{t-p}$ = Independent Variables that are dependent variables lagged specific time periods
- $\phi_0, \phi_1, \phi_2, \phi_p$ = Regression coefficients
- ϵ_t = Residual term that represents random events not explained by model

In this equation the regression coefficients are calculated by using the nonlinear least squares method. The nonlinear least squares method usually adapts an iterative solution approach to calculate the parameters rather than using a direct computation. This method starts by using the preliminary estimates and continues by systematically improving estimates until the optimal and most accurate values are identified. If the AR model captures successfully the dependence structure of the time series then the residuals should appear as randomly distributed. How the method exactly works is described mathematically in the third section. If the AR model captures successfully the dependence structure of the time series then the residuals should appear as randomly distributed. An AR signature in the distribution would be if the Partial autocorrelation function displays a sharp cut off while the ACF decays slowly with visible spikes at higher lags (*Hanke and Reitsch, 1998, P. 412-414).*

Figure 15: Autocorrelation and partial autocorrelation coefficients of AR(1) & AR(2) models



Source: Hanke, J. and Reitsch, A., 1998, Business Forecasting

Figure 15 represents the equations of an AR model of order one – AR(1) and an AR model of order two – AR(2). Number of *p* terms can be added to the equation in order to represent AR(*p*) model, where *p* is the number of past observation that are going to be included in the forecast for the upcoming period. The top two rolls of the figure represent the behaviour of the theoretical autocorrelation and partial autocorrelation functions for a simple AR(1) model. The main difference that can be spotted between the two figures is the behaviour of the correlations. The autocorrelation confidents progressively decline to zero, while the partial autocorrelation coefficients dopes to zero after the very first time lag. The second part of figure 10.2 *c)* and 10.2 *d)* show the behaviour of an AR(2) model. In this case, similarly to AR(1), the autocorrelation coefficients trail off to zero, while the partial autocorrelation coefficients drop to zero after the second time lag. This type of pattern will persist in any AR(p) model, with the only difference being in the case of partial autocorrelation. In this scenario the coefficient will drop to zero after the p time lag. However, it must be noted that sample autocorrelation functions are going to differ from these sampling functions due to sampling variation (*Hanke and Reitsch, 1998, P. 413).*

Moving-Average Models (MA)

The moving-average model takes the below form when put into an equation:

$$Y_t = w_0 + \epsilon_t - w_1\epsilon_{t-1} - w_2\epsilon_{t-2} - \cdots - w_q\epsilon_{t-q} \tag{52}$$

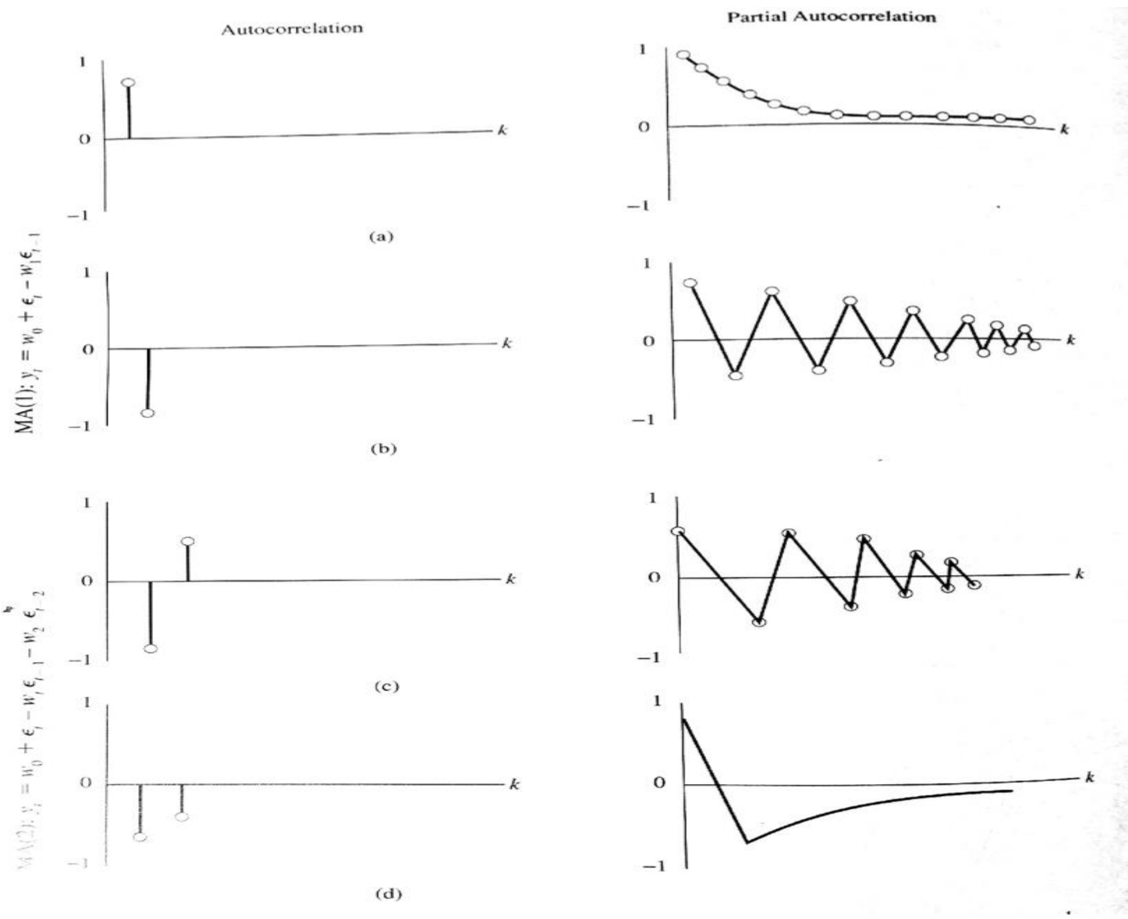
where

- Y_t = Dependent variable
- w_0, w_1, w_2, w_q = Weights
- ϵ_t = Residual or error
- $\epsilon_{t-1}, \epsilon_{t-2}, \epsilon_{t-q}$ = Previous values of residuals

The main difference between AR and MA are the independent variables used to calculate the dependable. As seen above in AR the dependent variable is calculated through the lagged values of the variable itself. In the case of an MA model, the variable depends on the previous values of the residuals. The variable is then estimated while using a linear combination of past errors. The weights are usually shown with negative coefficients although the weights can be both negative and positive. The sum of all the weights does not equal 1. The current values of the variable can be found from past shocks or errors added to new shocks or errors. The time series is regarded as a moving average that is unevenly weighted, because of different coefficient. If this model is accurate and captures the dependence structure in the data then the residuals should look random (*Hanke and Reitsch, 1998, P. 414*).

Figure 16 represents the equations of an MA model of order 1 MA(1) and MA model of order 2 MA(2). Number of *p* terms can be added to the equation in order to represent MA(*p*) model, where *p* is the number of past error terms that are going to be included in the forecast for the upcoming period. Figure 16 a) and b) illustrate the behaviour of the theoretical autocorrelation coefficients of the MA(1) model. One of the reasons to assess the correct model through the autocorrelation and the partial autocorrelation functions is the distinctive difference in their behaviour. The autocorrelation coefficients for the MA(1) model drops to zero after the first lag, while the partial autocorrelation trails off to zero in a very graduate manner. It is the same case concerning the model of MA(2). The only difference being that in the case of autocorrelation the coefficient drops to zero after the second lag. Here again, the partial autocorrelation function steadily decreases to zero. Here again it should be noted that sample autocorrelation functions would differ from these theoretical functions due to sampling variation (*Hanke and Reitsch, 1998, P. 414*).

Figure 16: Autocorrelation and partial autocorrelation coefficients of MA(1) and MA(2) models



Source: Hanke, J. and Reitsch, A., 1998, Business Forecasting

In order to summarise the above two models, it can be concluded that in the case of AR(1) model the autocorrelation coefficient function declines in geometric progression from its highest value at lag 1, while PACF cuts off abruptly after lag 1. In the case of the MA(1) model it is simply the other way around.

Autoregressive Moving-Average Models (ARIMA)

The ARIMA model is simply a mixture form AR and MA models and this is best represented by the equation below:

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \epsilon_t - w_1 \epsilon_{t-1} - w_2 \epsilon_{t-2} - \cdots - w_q \epsilon_{t-q} \tag{53}$$

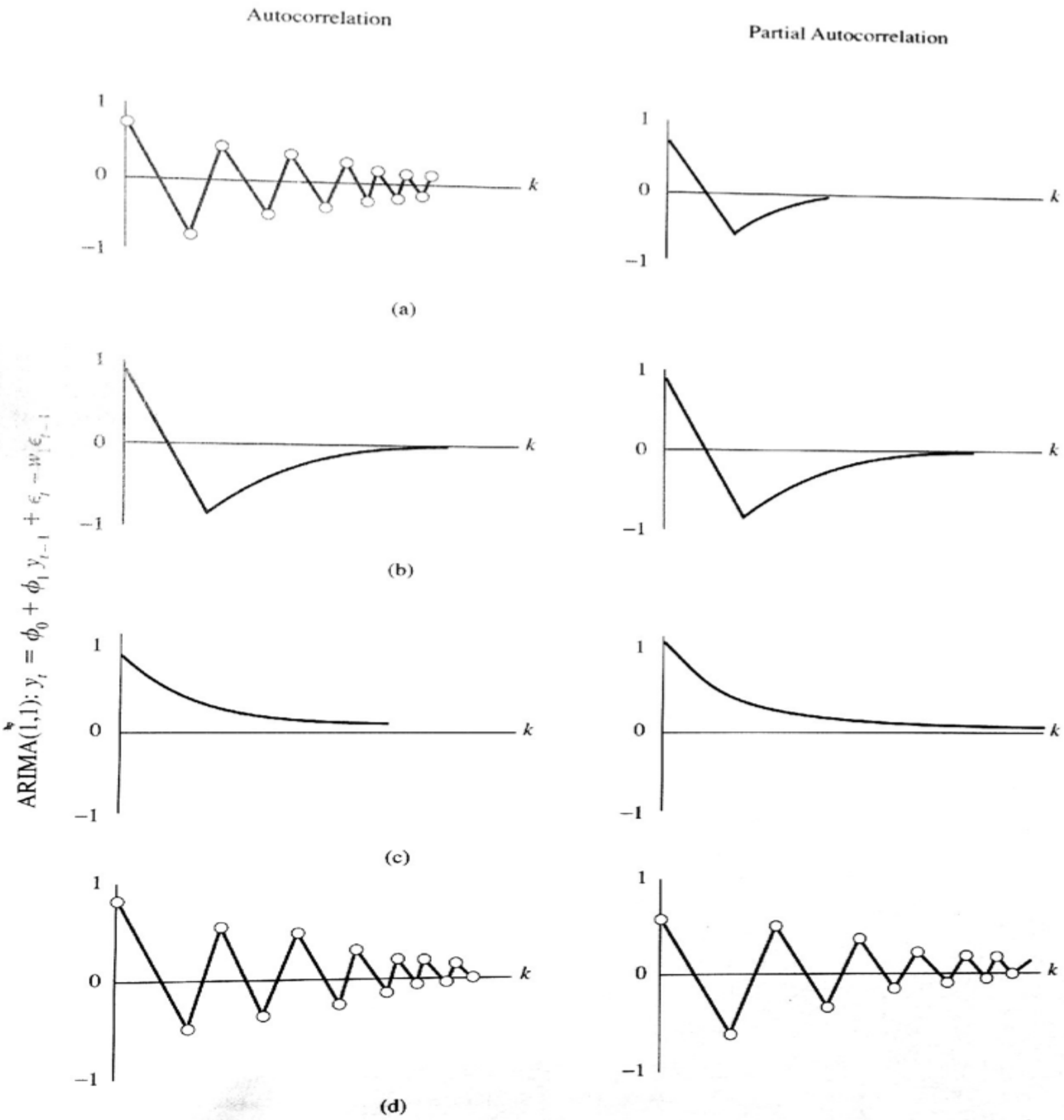
ARIMA models use a combination of past values and past errors and offer the potential of fitting models that could not be otherwise appropriately fitted using simply an AR or MA. In this case the letter "I" stands for Integrated, which means a differentiating must be done. One of the main characteristics of an ARIMA model is that it can be used for series with stochastic trends. Stochastic trend usually refers to random trends where they are not predicted by a certain event. This model is usually written as ARIMA (*p*,*d*,*q*), where:

- *p* is the number of autoregressive terms;
- *d* is the order of differencing and;
- *q* is the number of moving average terms.

Figure 17 shows the equation of an ARIMA(1,0,1), which is a first order AR model with no differencing and one moving average term. It represents the behaviour of the theoretical autocorrelation and partial autocorrelation coefficients (*Hanke and Reitsch, 1998, P. 415).*

A significant characteristic of Box-Jenkins methodology in comparison to other methods is the fact that it does not make assumptions about the number of terms or the relative weights to be assigned to the terms. It is the analyst who selects the appropriate model along with the number of terms that should be used. It is the program then that calculates the coefficients by using the nonlinear least squares method. The estimates that are provided are then put into confidence intervals (*Hanke and Reitsch, 1998, P. 415).*

Figure 17: Autocorrelation and partial autocorrelation coefficients of mixed ARIMA (1,1) models



Source: Hanke, J. and Reitsch, A., 1998, Business Forecasting

3.2 Applying the methodology

As previously discussed, the Box-Jenkins approach includes four separate stages:

1. Model construction
2. Model calibration and testing of model adequacy
3. Model validation

This work will now closely examine each of the three stages above.

Model construction

This first step of identifying the model is to determine whether the series is composed by stationary or non-stationary data. Stationary data is referred to the case when the mean value is not changing over time. If the series is not stationary the analyst should convert them to stationary through the method of differencing. He then has to specify the degree of differencing so the Box-Jenkins algorithm can convert the series into stationary data. Subsequent computation can then be undertaken using the converted data. Once the data is in the correct stationary format, the analyst can continue by identifying the most appropriate model. This is achieved by analysing and comparing the autocorrelation and the partial autocorrelation coefficients. This distribution and behaviour is closely analysed in order to be fitted in one of the examples discussed above. As mentioned, in some cases the data should be further manipulated in order to obtain clear results and identify it with one of the theoretical distributions above. It is expected that the analyst will be able to match the acquired corresponding coefficients to the distinctive feature of one of the models. In case the tests are inconclusive and the data cannot be properly matched, further tests regarding the accuracy of the model should be undertaken in Stage 2. The easiest way to identify the appropriate model is by studying the distribution. If the autocorrelation drops to zero exponentially it is regarding an AR model. However, if it is the partial autocorrelation that exponentially decreases to zero, it can be concluded that and MA model is required. Furthermore, if both correlations just trail off to zero then it can be assumed that an ARIMA model should be used. Finally, by counting the numbers of autocorrelation and partial autocorrelation coefficients that are statistically different from zero, the analyst will be able to determine the order of the MA and/or AR processed (*Hanke and Reitsch, 1998, P. 416).*

Model calibration and testing of model adequacy

After the model has been selected, the next step is to estimate the parameters. The Box-Jenkins computer program can do the calculations of the parameters such as the coefficients and the errors using the nonlinear least squares method. Then the analyst should check the accuracy of the model forecasts. This is achieved by examining the error terms $\epsilon_t = Y_t - \hat{Y}_t$ and to make sure they are random. This can be done by analysing the autocorrelations of the error terms and making sure they are not significantly different from zero. If there is a number of low-order or seasonal lags that are statistically different from zero, then the model should be revise and most probably considered as inadequate. Another way of checking the appropriateness of the model is through a chi-square test. This is also known as the modified Box-Pierce Q statistic, which is executed on the autocorrelations of the residuals. The test statistic is:

$$Q_m = n(n+2)\sum_{k=1}^{m}\frac{r_k^2}{n-k} \tag{54}$$

and is approximately distributed as a chi-square distribution with *m-p-q* degrees of freedom.

- n = number of observations in the time series
- k = the time lag to be checked
- m = the number of time lags to be tested
- $r(k)$ = Sample autocorrelation function of the k-th residual term

If the resulted Q value is larger than X squared for *m-p-q* degrees of freedom, then the analyst can conclude that the model is inadequate and return to the previous stage.

Those two tests should always be conducted in order to provide inside on the fitness of the model. In rare cases large deviations can be ignored. This is considered acceptable when those large deviations are clearly explained and caused by external factor that are unlikely to be repeated in the scope of the forecasts. Once the previous steps have been completed successfully and an adequate model has been chosen it can be proceeded to forecasting one or several periods into the future. (*Hanke and Reitsch, 1998, P.416- 418).*

Validating the model

Usually, as more data becomes available, the same model can be used to revise the previous forecasts by choosing another time origin. The analyst should always revise the new available data in order to identify any important changes that would require a change of the model. In such cases, where the series appears to be changing over time, the parameters might need to be recalculated. If the new parameters do not improve the model, a new one must be developed. The analyst should make this decision based on the difference in forecasts errors. If small ones are noticed, they may indicate that the parameters should be recalculated. When there are large differences in the forecast errors then the analyst should go back to the very beginning of the process where the data series is fitted into different models in order to identify the most suitable one (*Hanke and Reitsch, 1998, P. 418).*

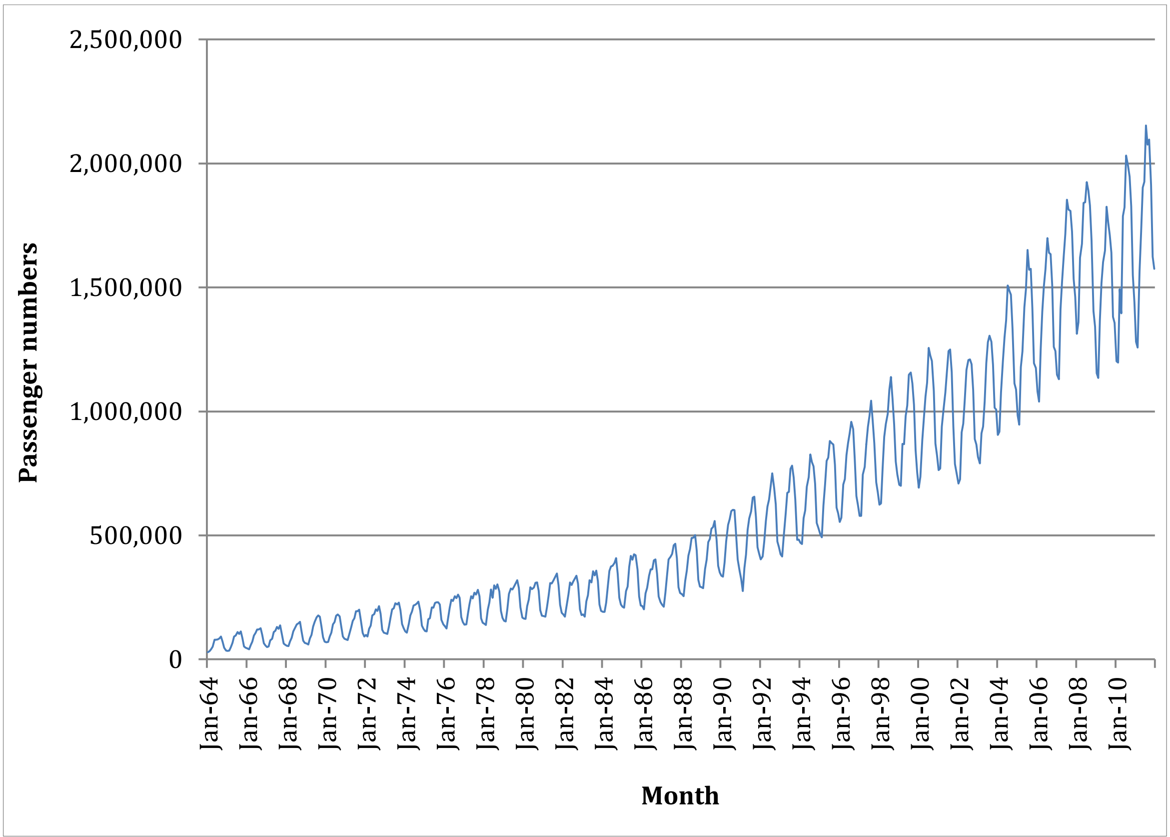
3.3 Vienna International Airport- Passenger demand forecasting

3.3.1 Model construction

After gaining understanding of the theory and techniques behind the ARIMA model, it can be now applied to the set of data in order to provide a one-year forecast for the passenger demand on the Vienna International Airport (VIE). The passenger demand between the years 1964 and 2011 will be used as the main historical data set. The numbers have been obtained from *Statistik Austria*.

The first step when forecasting, regardless of the methodology chosen is to plot the historical data and to analyse it gross before starting with the calculations. The original data$\{X_t\}_{t=1}^{576}$, is plotted in Figure 18. By observing the graphic it can be seen that the passenger demand on the Vienna International Airport is seasonal and that it has started to increase notably after the year 1990.

Figure 18: Original historical data of the passenger demand 1964-2011

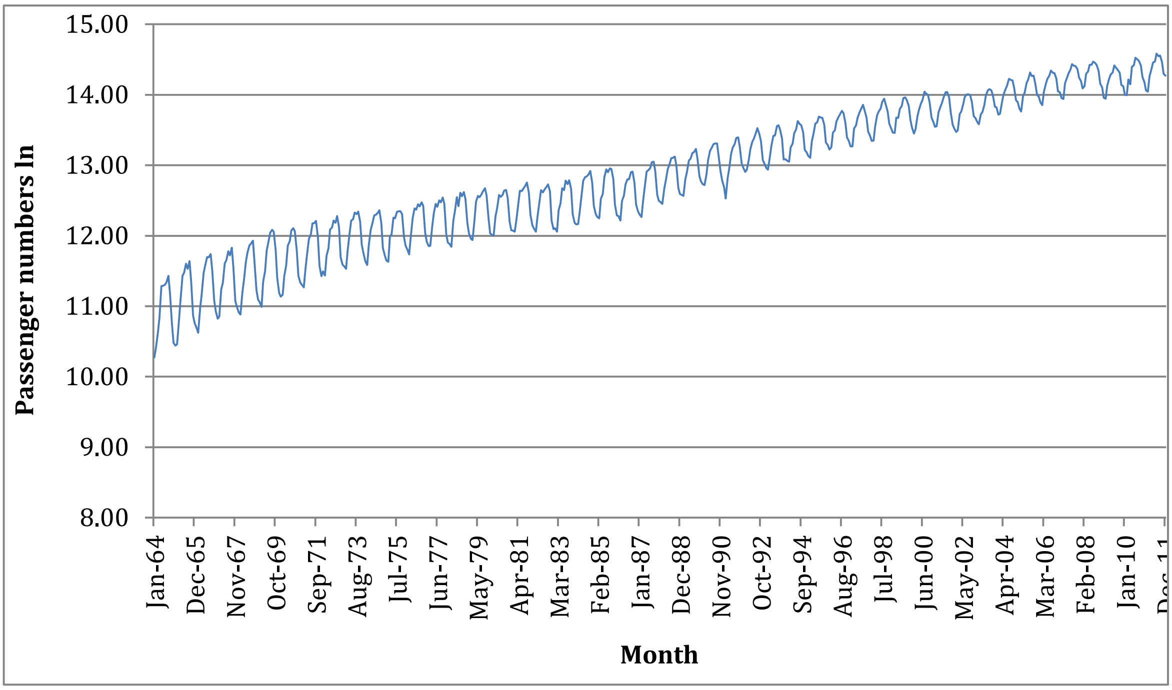


Because the fluctuations get broader over the time one has to make them as similar as possible, which means that the historical data must be logarithmized:

$$Y_t = \ln(X_t) \tag{55}$$

From now on one has to do the whole forecast with this logarithmized passenger numbers and at the end, when the prediction is completed, to de-logarithmize the outcomes in order to get the results in the same format as in the beginning of the project. The result from equation (55) is shown in Figure 19.

Figure 19: Logarithmized historical data of the passenger demand 1964-2011



By observing the graph it can be concluded that the trend from the original set of data persists in the logarithmized model as well. To prove this assumption mathematically, one has to analyse the autocorrelations of the data, something that is recommended regardless of how obvious the trend is. "The autocorrelations are the correlation between a variable, lagged one or more periods, and itself "(*Hanke and Reitsch, 1998, P. 91)*. They are used to identify time series data patterns and are calculated by the following formula:

$$\rho_k = \frac{\sum_{t=k+1}^{n}(Y_t - \bar{Y})(Y_{t-k} - \bar{Y})}{\sum_{t=1}^{n}(Y_t - \bar{Y})^2} \tag{56}$$

where:

- ρ_k = autocorrelation coefficient for a lag of k periods
- $\bar{Y}$ = mean of the values of the series
- Y_t = observation of time period t
- Y_{t-k} = observation k time periods earlier or at time period t-k

(*Hanke and Reitsch, 1998, P. 92).*

The autocorrelations are then tested for significance with the equations (57) and (58), which calculate the upper and lower limits. If these are exceeded, then the autocorrelation coefficients are said to be significantly different from zero.

$$Upper\ limit = t_{.975} \cdot SEr(k) = t_{.975} \cdot \sqrt{\frac{\left(1 + 2 \cdot \sum_{i=0}^{k-1} (\rho_i)^2\right)}{n}} \tag{57}$$

$$Lower\ limit = t_{.025} \cdot SEr(k) = t_{.025} \cdot \sqrt{\frac{\left(1 + 2 \cdot \sum_{i=0}^{k-1} (\rho_i)^2\right)}{n}} \tag{58}$$

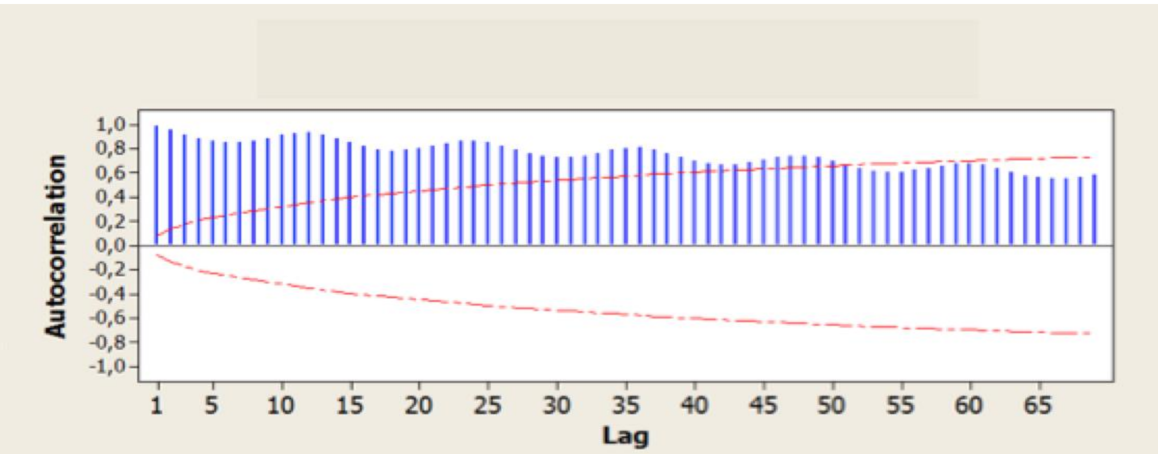
where

- $SEr(k)$ = standard error of the autocorrelation at lag k
- $\rho(i)$ = Autocorrelation at lag i
- n = number of observations in the data series

(*Hanke and Reitsch, 1998, P. 96-97).*

The results from equations (56), (57), (58) can be seen in Figure 20, where the blue columns describe the autocorrelations and the red dotted lines, the 95% confidence limits.

Figure 20: ACF for the logarithmized Passenger Demand



Lag	ACF	T	LBQ	Lag	ACF	T	LBQ
1	0,980969	23,54	557,18	6	0,847710	6,62	2871,73
2	0,950988	13,35	1081,73	7	0,849514	6,18	3293,99
3	0,915531	10,10	1568,74	8	0,863214	5,90	3730,74
4	0,882083	8,36	2021,61	9	0,886379	5,72	4192,07
5	0,859153	7,31	2452,00	10	0,909984	5,57	4679,15

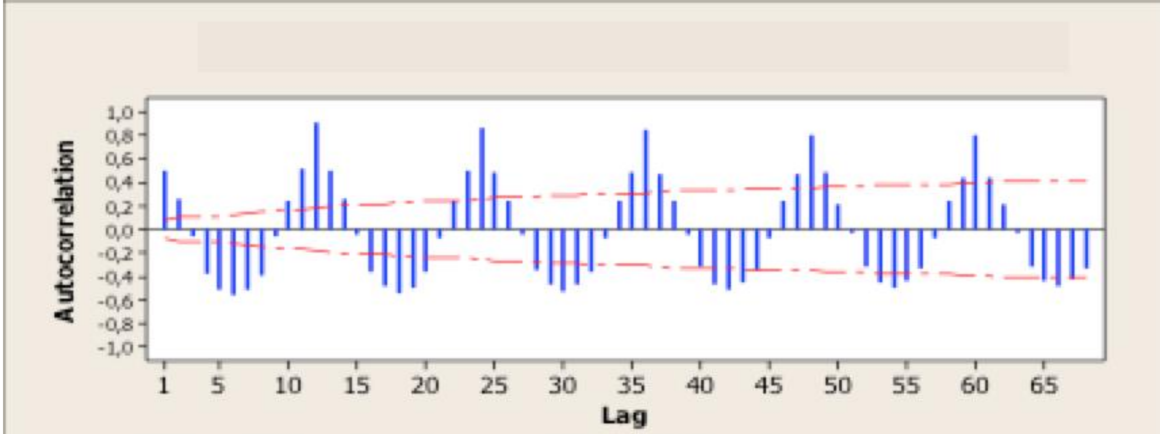
By looking at the autocorrelations one can see that for the first fifty periods the autocorrelation coefficients are significantly different from zero, they are above the red limits line, and then with the increasing of the periods they decrease to zero. This confirms

that the original observation is true and that the logarithmized historical data did have a trend and due to the fact that the average value is changing over time it is not stationary.
As described in the methodology part, in order to eliminate the trend one has to difference the data. The first-difference is calculated as follows:

$$\nabla Y_t = Y_t - Y_{t-1}. \tag{59}$$

The plot of the autocorrelations of the first-difference, displayed in Figure 21, shows that the trend is removed, thus a stationarity have been achieved on a period-to-period basis. However the autocorrelation coefficient for the time lags 12, 24, 36, etc. are very high, which signals that our observation done in the beginning is true. The data is periodical with the period 12 (a seasonal component is available in the series ∇Y_t), which means that the series is not stationary yet. If the autocorrelation coefficient for lag 24 has dropped to zero, then the first-differenced data would have been stationary.

Figure 21: ACF for the logarithmized Passenger demand: First differenced



Lag	ACF	T	LBQ	Lag	ACF	T	LBQ
1	0,499636	11,98	144,29	6	-0,565779	-8,70	602,21
2	0,260965	5,11	183,72	7	-0,513164	-7,02	756,03
3	-0,057150	-1,07	185,62	8	-0,386914	-4,89	843,63
4	-0,371438	-6,95	265,78	9	-0,060043	-0,73	845,74
5	-0,507271	-8,78	415,56	10	0,244370	2,96	880,80

By using the long-term difference (differences with a length of L periods) one can eliminate the trend for lags 12, 24, 36, etc. The long-term differences are mathematically defined in equation (60) on the next page.

$$\nabla\nabla_{12}Y_t = (Y_t - Y_{t-L}) - (Y_{t-1} - Y_{t-L-1})$$
$$= (Y_t - Y_{t-1}) - (Y_{t-L} - Y_{t-L-1}) \qquad (60)$$

where

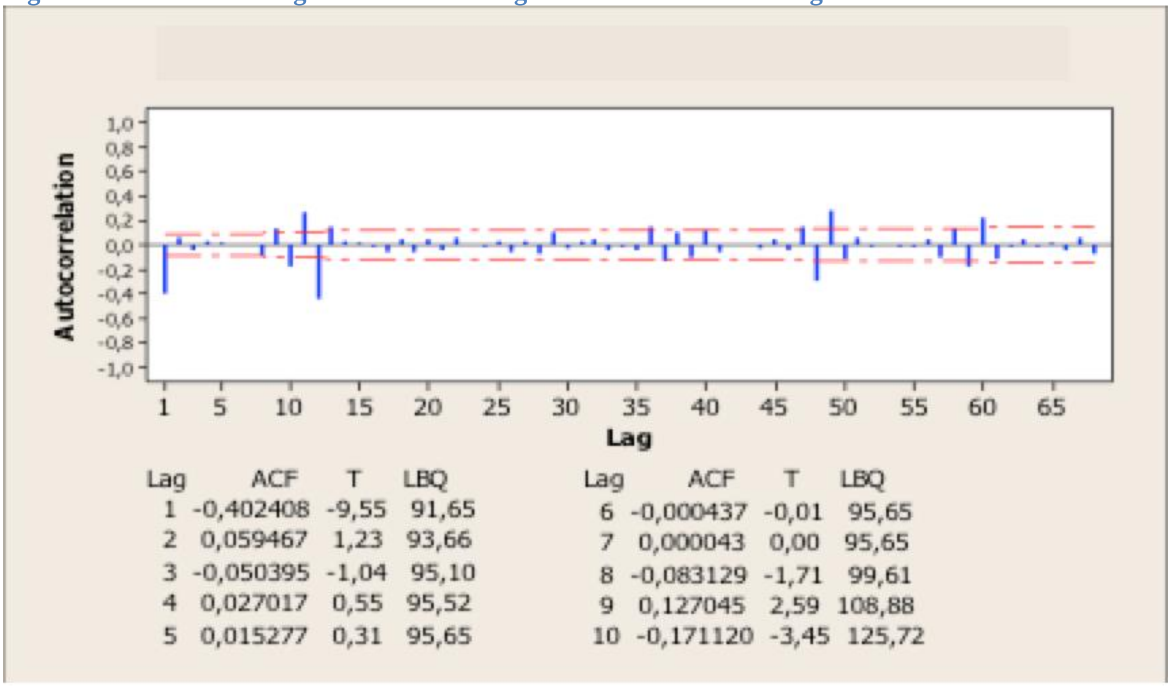
1. L =12

(*Hanke and Reitsch, 1998, P. 447*).

Please note that the methodology of long differencing is the same for every periodicity that can be found in a series. For example if the autocorrelations were significantly different from zero for the months 4, 8, 12, etc. then one would have got a periodical data with the period 4 and in the long-term difference calculation equation L would be standing for 4 periods.

One can see in Figure 22 that the first and long-differenced data is stationary because as good as only the autocorrelations for lags 1 and twelve are significantly different from zero and the rest is randomly dispersed around zero.

Figure 22: ACF for the logarithmized Passenger Demand: short- & long-differenced



Having achieved a stationarity in the series one can continue to the second step of the first stage described in section 3.2- finding an appropriate model. In order to do this one has to analyse both the autocorrelation function and the partial autocorrelation function for the short- and long-differenced data.

The partial autocorrelation describes the dependence of Y_t and Y_{t-k}, when the dependence on all other variables $Y_{t-1}, Y_{t-2}, \cdots, Y_{t-k-1}$are removed, e.g. Z is regressed upon $Y_1$and Y_2 then it is of interest to ask how much explanatory power Y_t has if the effect of Y_2 is partielled out (C*ryer and Chan, 2008, P. 112).*

For any stationary process the partial autocorrelation function can be calculated using the following general method:

$$\begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{p-1} \\ \rho_1 & \ddots & & & \\ \rho_2 & & \ddots & & \\ \vdots & & & \ddots & \\ \rho_{p-1} & \rho_{p-2} & \rho_{p-3} & \cdots & 1 \end{bmatrix} \begin{bmatrix} \phi_{k1} \\ \phi_{k2} \\ \phi_{k3} \\ \vdots \\ \phi_{kp} \end{bmatrix} = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \rho_3 \\ \vdots \\ \rho_p \end{bmatrix} \tag{61}$$

that equals

$$\begin{matrix} \phi_{k1} & + & \rho_1\phi_{k2} & +\cdots+ & \rho_{k-1}\phi_{kk} & = & \rho_1 \\ \rho_1\phi_{k1} & + & \phi_{k2} & +\cdots+ & \rho_{k-2}\phi_{kk} & = & \rho_2 \\ \vdots & & \vdots & \cdots & \vdots & = & \\ \rho_{k-1}\phi_{k1} & + & \rho_{k-2}\phi_{k2} & +\cdots+ & \phi_{kk} & = & \rho_k \end{matrix} \tag{62}$$

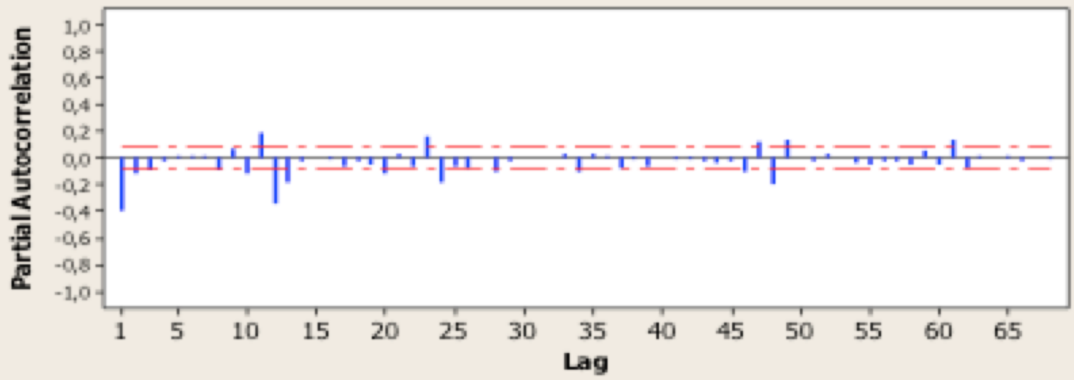
where:

- ρ_p = the autocorrelation
- ϕ_{kk}= partial autocorrelation (not to be mistaken for a parameter of an AR model)

(C*ryer and Chan, 2008, P. 113-114*)

A graph of the partial autocorrelation function is shown in Figure 23 on the next page.

Figure 23: PACF for the logarithmized Passenger Demand: short- & long-differenced



Lag	PACF	T	Lag	PACF	T
1	-0,402408	-9,55	6	0,017669	0,42
2	-0,122264	-2,90	7	0,011867	0,28
3	-0,088112	-2,09	8	-0,093553	-2,22
4	-0,026986	-0,64	9	0,065154	1,55
5	0,017798	0,42	10	-0,121486	-2,88

As previously discussed one can find the best fitting model examining the autocorrelations in Figure 22 and the partial autocorrelation in Figure 23, something that is done in two steps. The first one is to examine the non-seasonal pattern. Figure 22 shows that the autocorrelation coefficients drop of to zero after the first significant one, which equals -0,402 and Figure 23 shows that the partial correlation of the nonseasonal components is exponentially decaying to zero. When comparing those two with the graphs shown in 3.2 one sees that this two figures are indicating an IMA (1) model. The second step of identifying the model is to analyse the seasonal pattern. Both figures show again that an IMA (1) model would best fit. This is so because on the one hand the autocorrelation drops off to zero after the first significant coefficient at time lag 12 and on the other the seasonal partial autocorrelation shows again an exponential decay (see the coefficients for time lag 12, 24, 36).

3.3.2 Model calibration and testing of the model adequacy

After having selected a tentative model one can continue to Stage 2. Firstly one has to estimate the parameters, $w_1, W_1, \epsilon_{t-1}, \epsilon_{t-L}, \epsilon_{t-L-1}$, of the model. This can be done in two ways. The first one is by using maximum likelihood estimation and the second option is using the non-linear least squares method. As the software Minitab, which was used for the calculation of the parameters, uses the nonlinear squares method, a closer look at this estimation is given on the following pages.

Consider a set of m data points $(x_1, y_1), (x_2, y_2), \ldots , (x_m, y_m)$, and a model function $y = f(x, \beta)$, that depends on both the variable x and the parameters $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots , \beta_k)$, which are used as a general symbol for the k=p+q parameters $(\boldsymbol{\phi}, \boldsymbol{w})$, with $m \geq k$. One is willing to find the parameters of the vector $\boldsymbol{\beta}$ such that the curve fits best the given data in the least square sense. Mathematically this means that the sum of squares shown in the following equation should be minimized:

$$S = \sum_{i=1}^{m} \epsilon_i^2 \tag{63}$$

In equation (64) ϵ_i stands for the residuals, which are given by:

$$\epsilon_i = y_i - f(x_i, \boldsymbol{\beta}) \tag{64}$$

for $i = 1,2,3 \ldots m$.

The minimum value of S occurs when the gradient is zero and since the model contains n parameters there are n gradient equations:

$$\frac{\partial S}{\partial \beta_j} = 2 \sum_i \epsilon_i \frac{\partial \epsilon_i}{\partial \beta_j} = 0 \tag{65}$$

where j=1,…,n.

The derivatives $\frac{\partial \epsilon_i}{\partial \beta_j}$ are functions of both the independent variable and the parameters, in a non-linear system, so these gradient equation do not have a closed solution. Instead initial values must be chosen for the parameters. Then, the parameters are obtained by successive approximation:

$$\beta_j \approx \beta_j^{k+1} = \beta_j^k + \Delta\beta_j \tag{66}$$

At each iteration the model is linearized by approximation to a first-order Taylor series expansion about $\boldsymbol{\beta}^k$:

$$f(x_i, \boldsymbol{\beta}) \approx f(x_i, \boldsymbol{\beta}^k) + \sum_j J_{ij} \Delta\beta_j \tag{67}$$

where $J_{ij} = -\frac{\partial \epsilon_i}{\partial \beta_j}$. As the Jacobian J is a function of constants, the independent variable and the parameters it changes for the different iterations. By adopting equation (67) in (64) one gets the following equation for the residuals:

$$\epsilon_i = \Delta y_i - f(x_i, \boldsymbol{\beta}) = y_i - f(x_i, \boldsymbol{\beta}^k) - \sum_{s=1}^{n} J_{is}\, \Delta\beta_s \tag{68}$$

By substituting these expressions into the gradient equation, re-arranging them to normal equations and writing them in matrix notation, one gets the following equation:

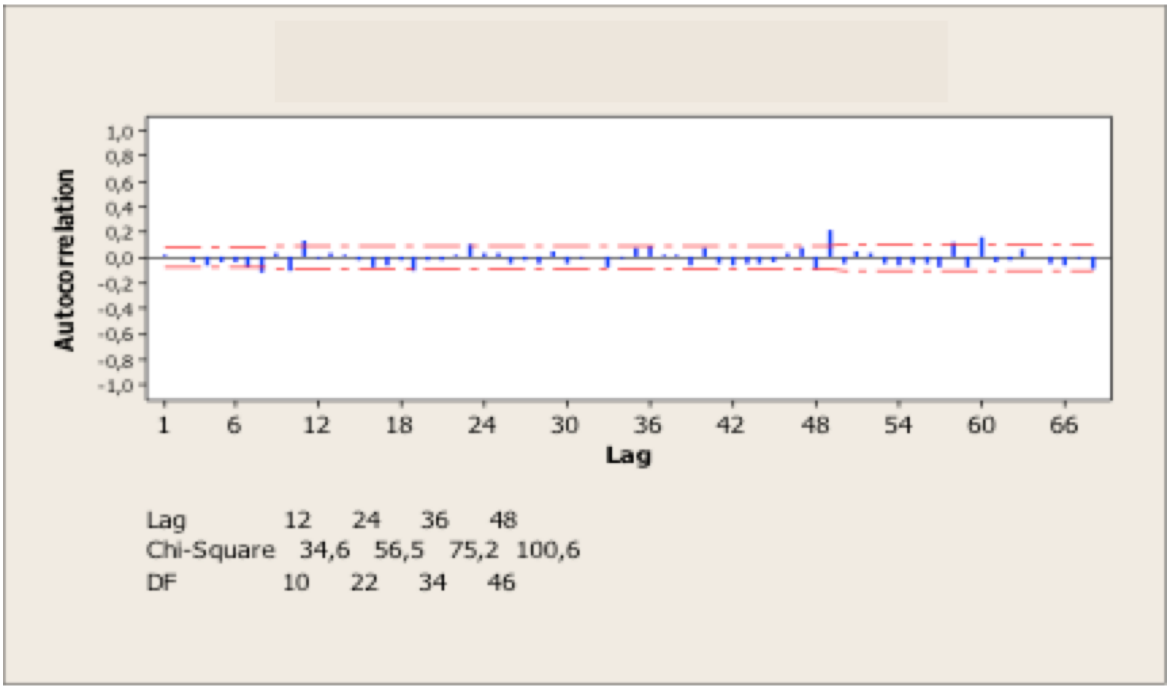
$$\Delta\beta = (J^T J)^{-1} J^T \Delta y \tag{69}$$

(*Box, Jenkins and Reinsel, 2008, P.255-256)*

Please note that the model was ran without a constant term, which is the approach when the differences are modelled. The parameters for the ARIMA (0,1,1)(0,1,1) model are estimated to be 0,4032 and 0,5838.

The residual autocorrelation coefficients can be seen in Figure 24.

Figure 24: ACF for the residuals of an ARIMA (0,1,1)(0,1,1) model



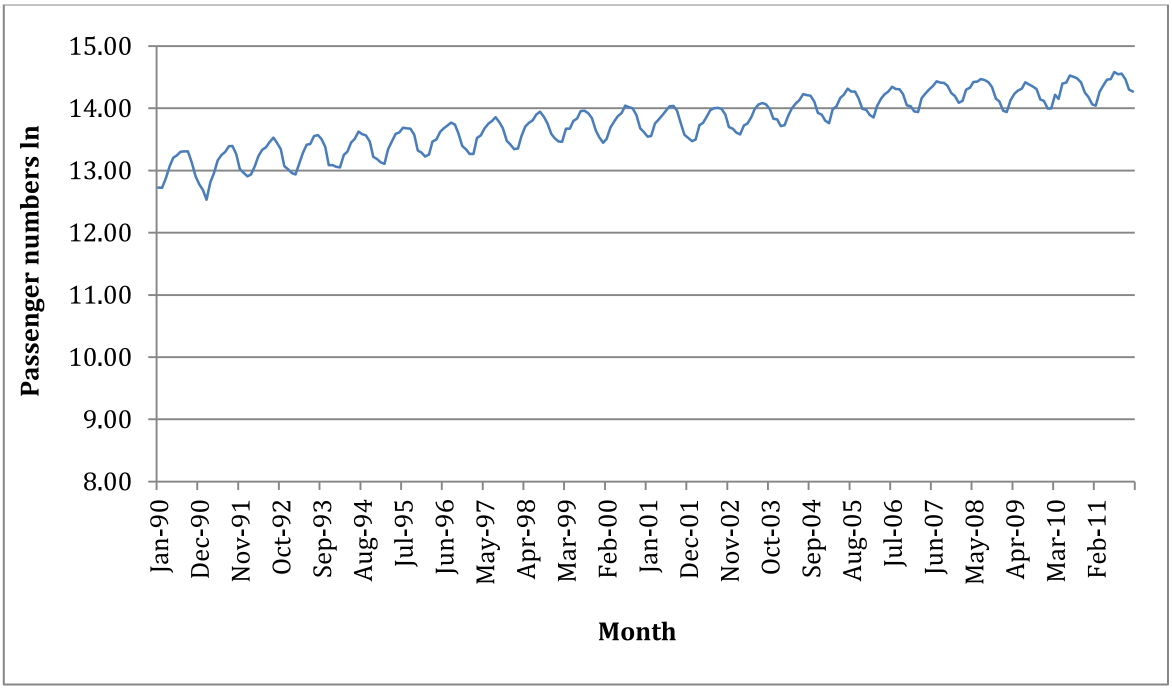
One can see that there are some of the low order coefficients that are significantly different from zero, which is normally a sign that the model is inadequate. In order to prove if the model is adequate a check with the modified Box-Pierce chi-square statistic must be made. The results from the test can be seen in Figure 24 under the graph and one can see that for all four degrees of freedom the chi-square statistics are higher than the tabulated values, shown in Figure 35 in the Appendix, at the .05 significance level. This is a sign that the model is inadequate and one has to go back to stage 1.

After taking a closer look at the historical data in Figure 18 one can identify the reason behind the inadequacy of the model. It is the fact that two different main trends can be found. The one before 1990, which appears to be very flat and the one afterwards, where the passenger numbers are increasing almost exponentially. This exponential increase is a result of many and different factors, but there is one that have influenced it most- the prices. The decreasing costs of flying in the last couple of decades and the growing availability has made flying more affordable, which leads to the fact that many people now a days prefer to fly than to use other types of transport for both long and short distances.

Although the rule is to use as many historical data as possible the two different trends are a sign that one has to design the model only using the historical data after the year 1989.

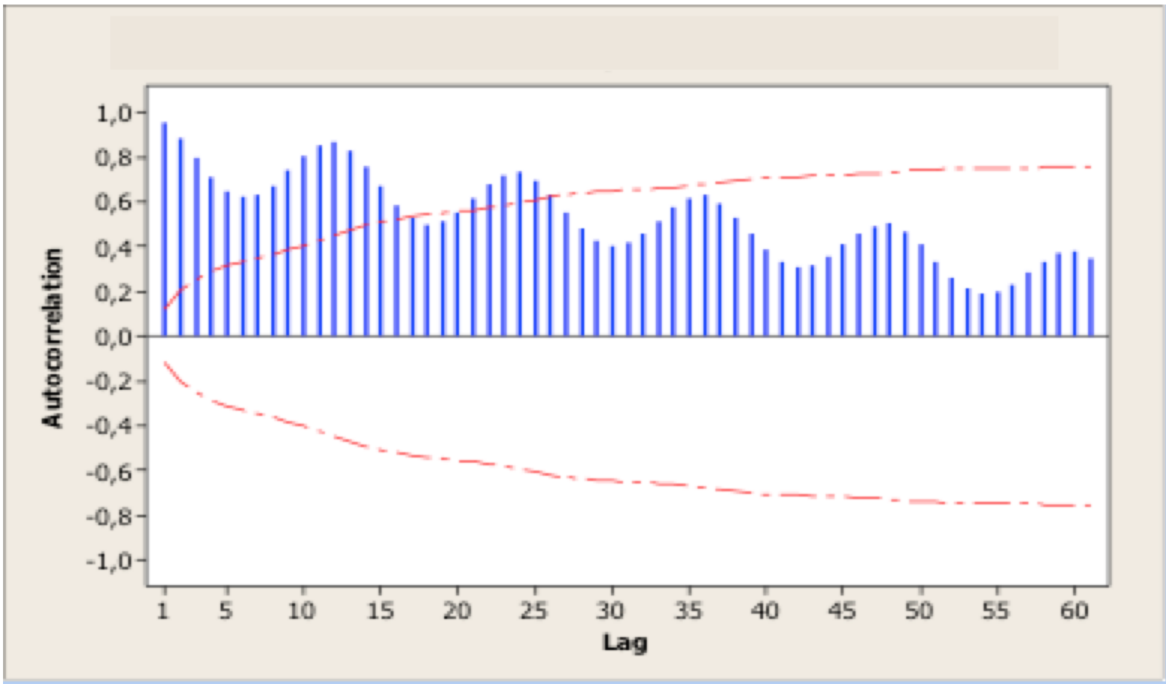
The logarithmized passenger numbers historical data for the years 1990- 2011 can be seen in Figure 25. Again one can recognize that there is a trend and seasonality in the series, but as mentioned above one has to investigate this by drawing the autocorrelation function.

Figure 25: Logarithmized historical data of the passenger demand 1990-2011



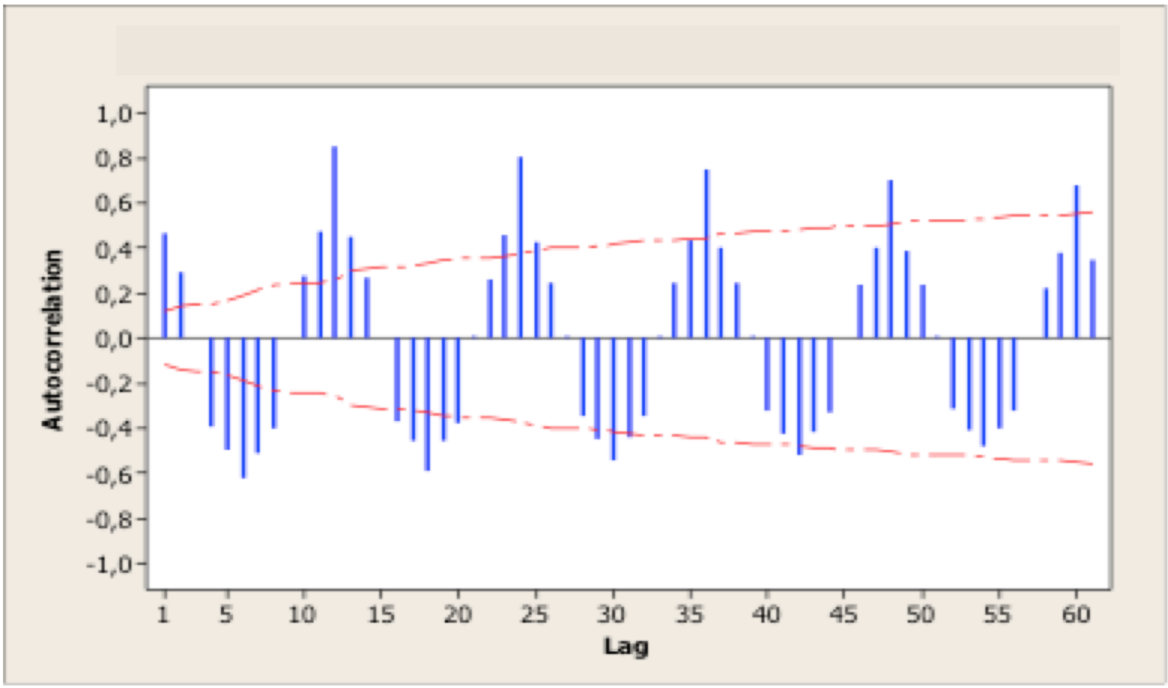
In Figure 26, where the autocorrelation coefficients for the logarithmized data between 1990 and 2011 are shown, the autocorrelations at the beginning are above the levels and then decays to zero, which, as in Figure 20, is a sign for the expected trend.

Figure 26: ACF for the logarithmized Passenger Demand 1990-2011



As can be seen above, when a trend is identified, then the series must be differenced. This is done here as well and the autocorrelations of this short-differenced data can be seen in Figure 27. Here the expectations that there is a seasonal factor in the data is proven by the coefficients at lags 12, 24, etc. which are significantly different from zero. This means that again, one has to long-difference the data in order to eliminate the seasonal factor and to get a stationary data.

Figure 27: ACF for the logarithmized Passenger demand 1990-2011: First difference



In Figure 28 and Figure 29 one can see a plot of the functions that are needed in order to choose the appropriate model- the autocorrelation and the partial autocorrelation functions of the long-term differenced data. There is evidence that stationary data is obtained by the autocorrelations after lag 1 - the coefficients drop off to zero and that from the seasonal components only lag 12 is significantly different from zero.

Figure 28: ACF for the logarithmized Passenger Demand 1990-2011: short- & long-differenced

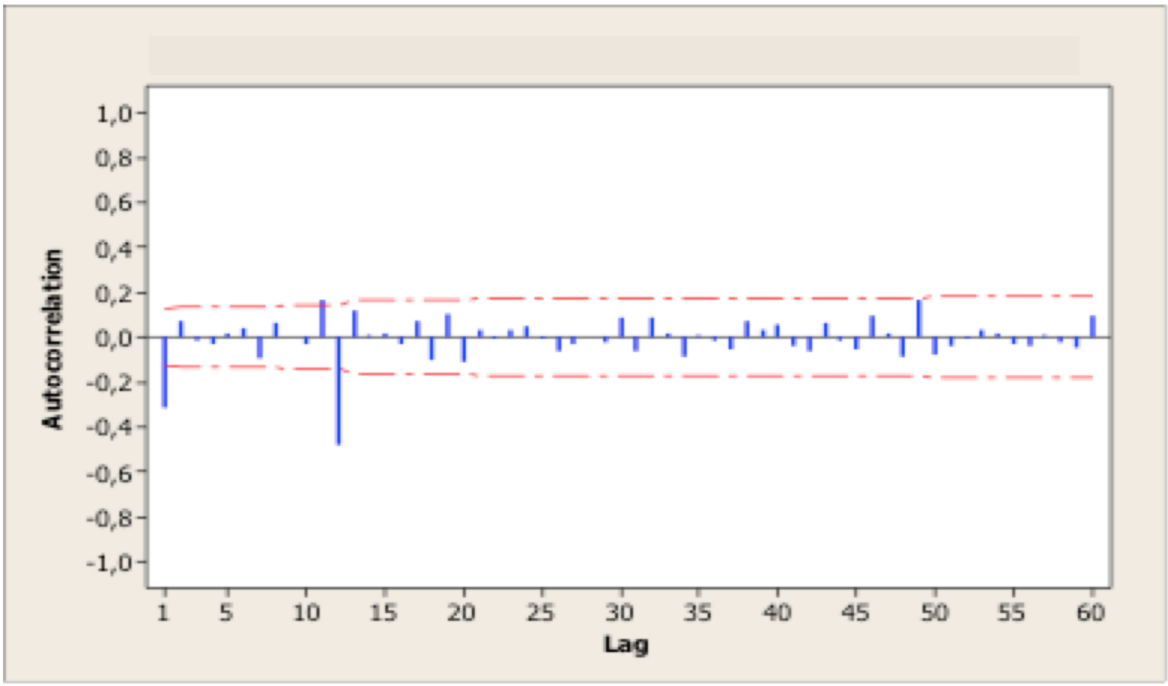
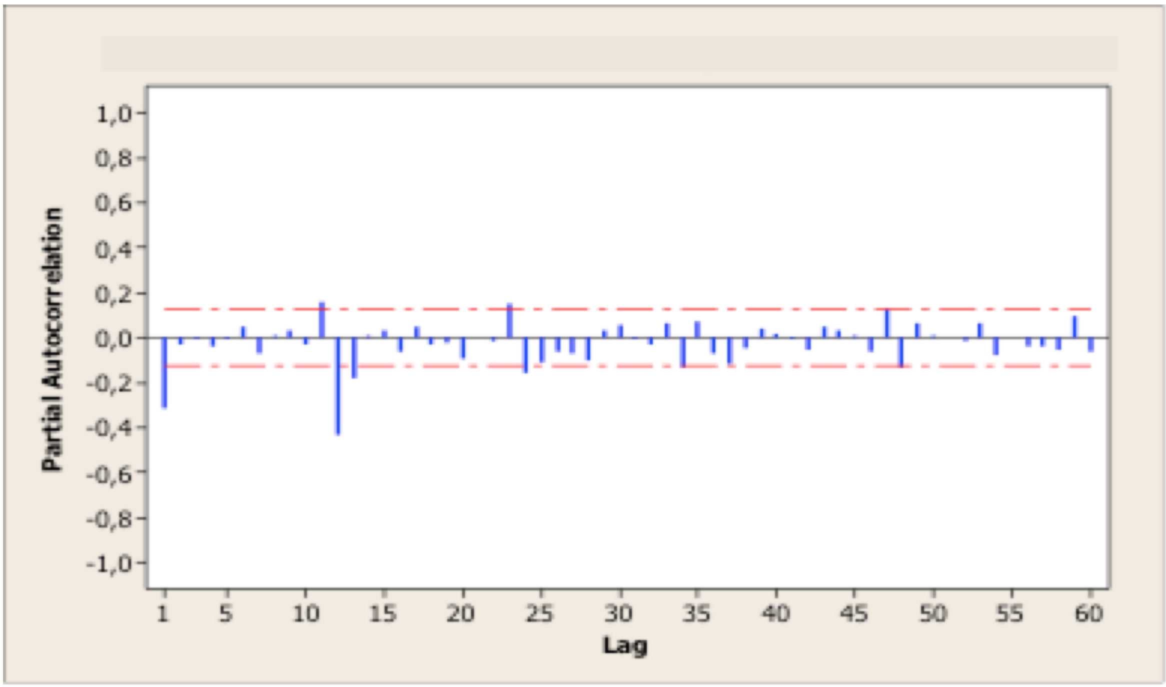


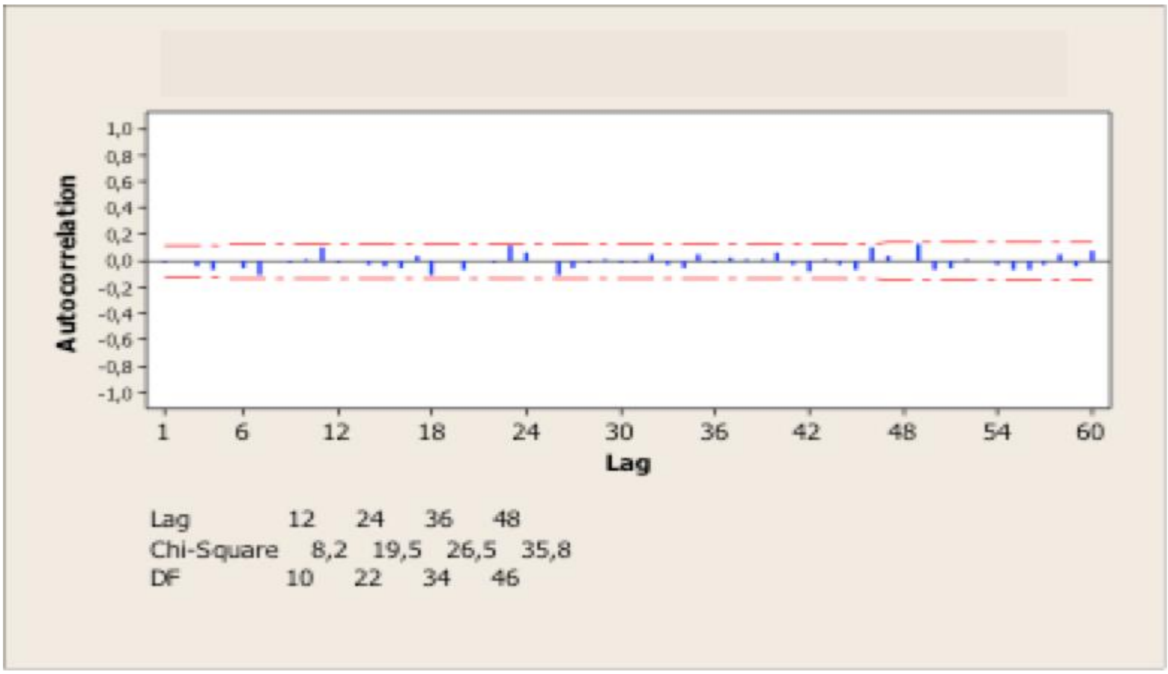
Figure 29: PACF for the logarithmized Passenger Demand 1999-2011: short- & long-differenced



As discussed above, when the whole available historical data was used one has to examine the nonseasonal and the seasonal patterns. By observing the graphs of autocorrelations one can see that only the first one is significant and that the partial autocorrelation coefficients exponentially drop off to zero. This is a sign that an IMA(1) process is the perfect fit. Since the same pattern can be found by seasonal coefficients one can consider that an IMA(1) is here also the best fit. Again the appropriate model is ARIMA (0,1,1)(0,1,1).

The parameters for the model were estimated to be 0,2655 and 0,7342 and the residuals autocorrelation function was again plotted in Figure 30, which is shown on the next page. One can see that in comparison to the ACF of the residuals from Figure 24 this time there are no low order coefficients that are significantly different from zero, which shows that the model is appropriate. This observation was confirmed by testing the chi-square statistics for 48 time lags, those results can be found in Figure 30. The calculated chi-squares are less then the tabulated values at the .05 significance level and therefore one can continue to Stage 3 and make a forecast with the model.

Figure 30: ACF for the residuals of an ARIMA (0,1,1)(0,1,1) model



Please note that at this point the data used is still the logarithmized one. The numbers should be de-logarithmized not before the forecast is made. For the forecast a multiplicative model was used to apply the nonseasonal model to the terms from a purely seasonal model. In the following equation (70) one can see the seasonal MA model written:

$$Y_t = u_t - W_1 u_{t-L} \tag{70}$$

where

- u_t = Residual or error
- W_t = Seasonal weights

When one substitutes for u_t the nonseasonal MA term applied to the errors

$$u_t = \epsilon_t - w_1 \epsilon_{t-1} \tag{71}$$

one gets the following equation:

$$\begin{gathered} Y_t = (\epsilon_t - w_1 \epsilon_{t-1}) - W_1(\epsilon_{t-L} - w_1 \epsilon_{t-L-1}) \\ Y_t = \epsilon_t - w_1 \epsilon_{t-1} - W_1 \epsilon_{t-L} + w_1 W_1 \epsilon_{t-L-1} \end{gathered} \tag{72}$$

(*Hanke and Reitsch, 1998, P. 448).*

Now the forecast for the period 265 (January 2012) is made using the ARIMA (0,1,1)(0,1,1) model, which was proven to be accurate by the chi-square statistic used above. Note that the differenced series is used for the forecast.

$$Y_t - Y_{t-1} - Y_{t-L} + Y_{t-L-1} = \epsilon_t - w_1\epsilon_{t-1} - W_1\epsilon_{t-L} + w_1W_1\epsilon_{t-L-1}$$
$$Y_t = Y_{t-1} + Y_{t-L} - Y_{t-L-1} + \epsilon_t - w_1\epsilon_{t-1} - W_1\epsilon_{t-L} + w_1W_1\epsilon_{t-L-1} \quad (73)$$

The forecast for period 265 is calculated as follows:

$$\hat{Y}_{265} = Y_{264} + Y_{253} - Y_{252} - w_1\epsilon_{264} - W_1\epsilon_{253} + w_1W_1\epsilon_{252}$$

$$\hat{Y}_{265} = 14{,}2695 + 14{,}0634 - 14{,}18 - 0{,}2655 \cdot 0{,}0121 - 0{,}7342 \cdot (-0{,}0179) + 0{,}2655 \cdot 0{,}7342 \cdot (-0{,}04) \quad (74)$$

$$\hat{Y}_{265} = 14{,}1550$$

The forecast for period 266 is:

$$\hat{Y}_{266} = \hat{Y}_{265} + Y_{254} - Y_{253} - W_1\epsilon_{254} + w_1W_1\epsilon_{253}$$
$$\hat{Y}_{266} = 14{,}1550 + 14{,}0449 - 14{,}0634 - 0{,}7342 \cdot (-0{,}0179) + 0{,}2655 \cdot 0{,}7342 \cdot (-0{,}0179) \quad (75)$$
$$\hat{Y}_{266} = 14{,}1462$$

One can see that the term $w_1\epsilon_{264}$ has not been updated to $w_1\epsilon_{265}$ but has disappeared because it equals to zero. This is due to the fact that all residuals after the time point 264 are equal to zero because there is no historical data to which the forecasted values can be compared and by the calculation of the residuals one subtracts the forecasted value from the forecasted value. The same method is applied to the last two terms in the equation if one is calculating the prediction for longer than one year.
The rest of the values can be found in the following Table 16:

Table 16: Forecasted values for the year 2012 using the ARIMA (0,1,1)(0,1,1) model

	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12
Y	14,1550	14,1462	14,3505	14,4048	14,5219	14,5507	14,6444	14,6186	14,6056	14,5296	14,3618	14,3225

Now having made the forecast using the logarithmized data, which prevents having fluctuations growing over the years, one can make the last step, where all the values are de-logarithmized. This is mathematically defined the following formula:

$$X_t = e^{Y_t} \tag{76}$$

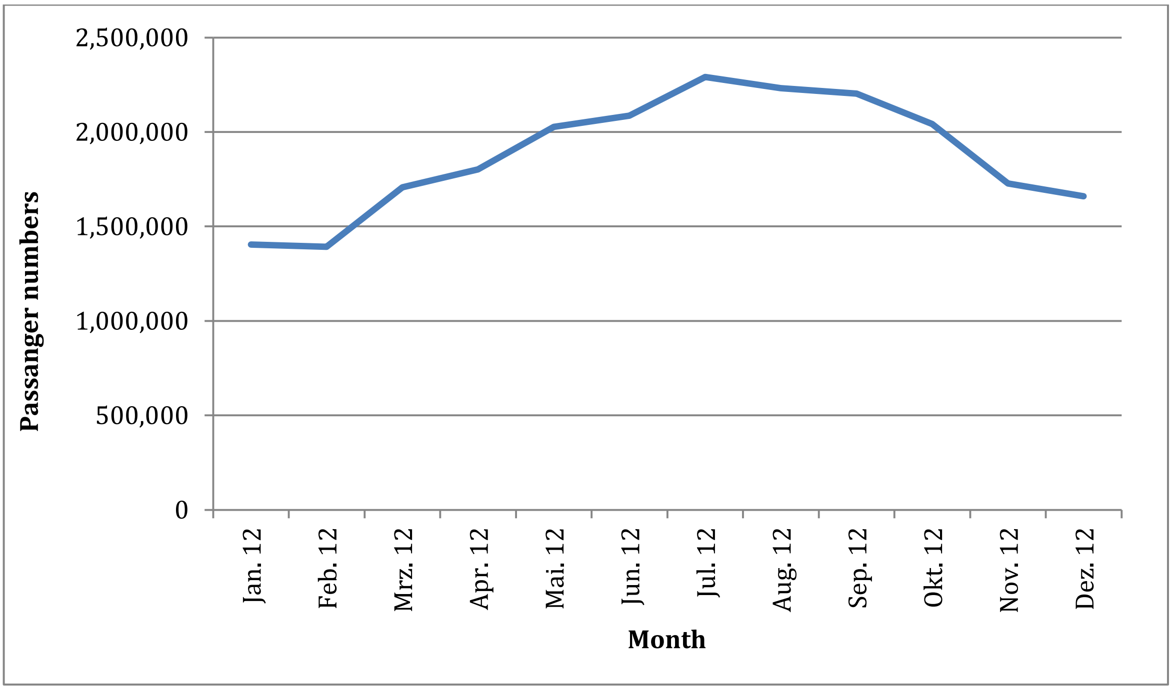
For the passenger numbers for the year 2012 one gets the following values:

Table 17: Forecasted de-logarithmized values for the year 2012 using the ARIMA (0,1,1)(0,1,1) model

	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12
Y	1.404.218	1.404.218	1.404.218	1.404.218	1.404.218	1.404.218	1.404.218	1.404.218	1.404.218	1.404.218	1.404.218	1.404.218

The plot of the forecasted values is given Figure 31.

Figure 31: Forecast ARIMA (0,1,1)(0,1,1)

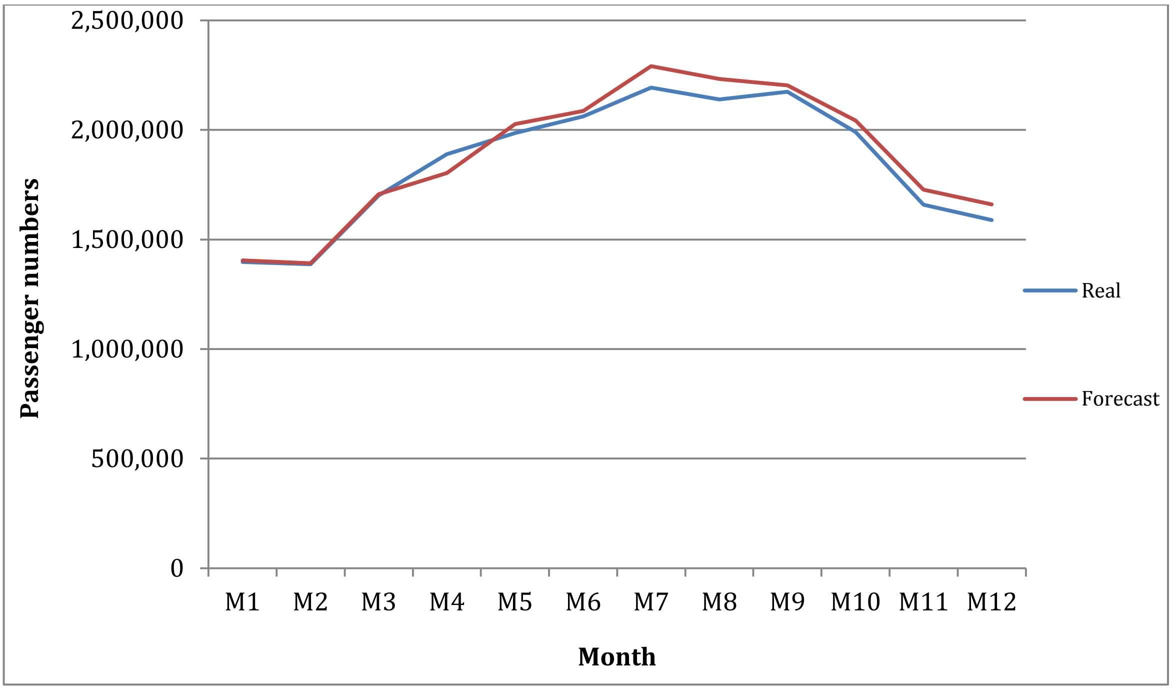


3.4 Model validation with out of sample tests

As in chapter 2 where the forecast was done using the Gaussian Enterprise Model, in this part again an out of sample test will be made to check if one has managed to forecast the future accurately by the chosen model. The difference this time is that a one-year, a four-year and a five-year forecast will be tested. The idea to also make a prediction for a longer period of time is to see how accurate it will be, because often it is important for airports to make long-term predictions in order to know if and when they will need a new runway, parking places, terminals, etc.

The plot of the one-year forecast compared to the real passenger numbers for year 2012 can be seen in Figure 32.

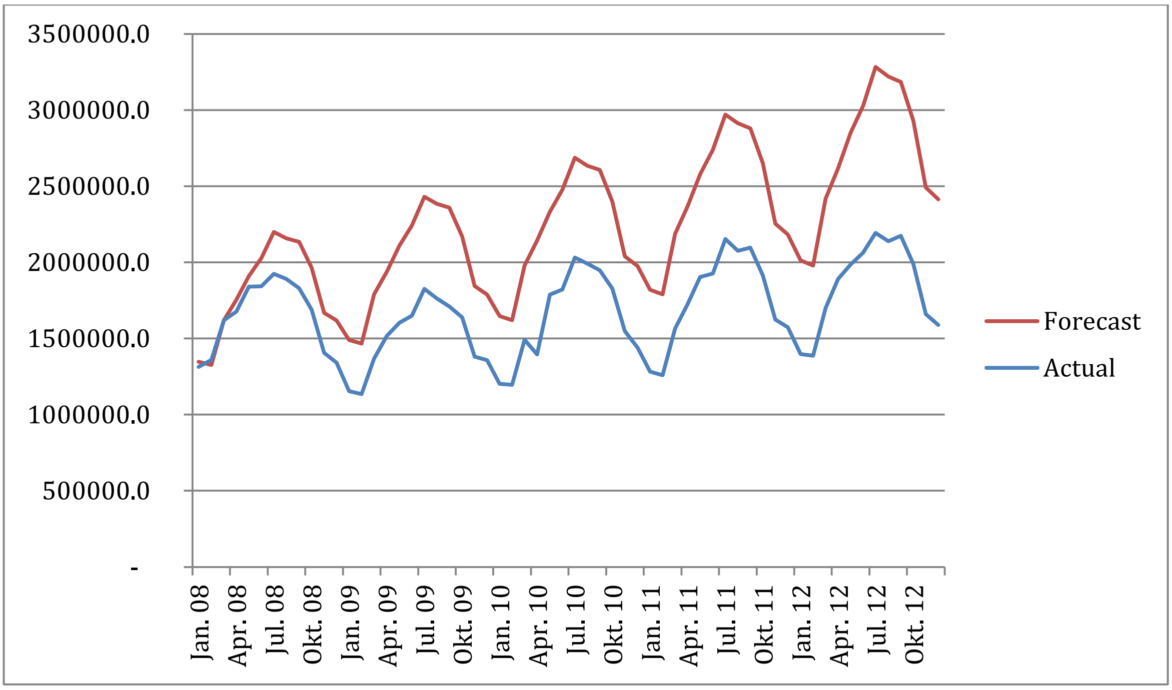
Figure 32: Out of sample test for the one-year forecast done by the ARIMA (0,1,1)(0,1,1) model



One can see on the plot that the red line, describing the forecasted values, is almost equal to the blue one, which stands for the real passenger numbers taken from the historical data. This is a clear sign that the methodology and the chosen model can describe the future very accurate.

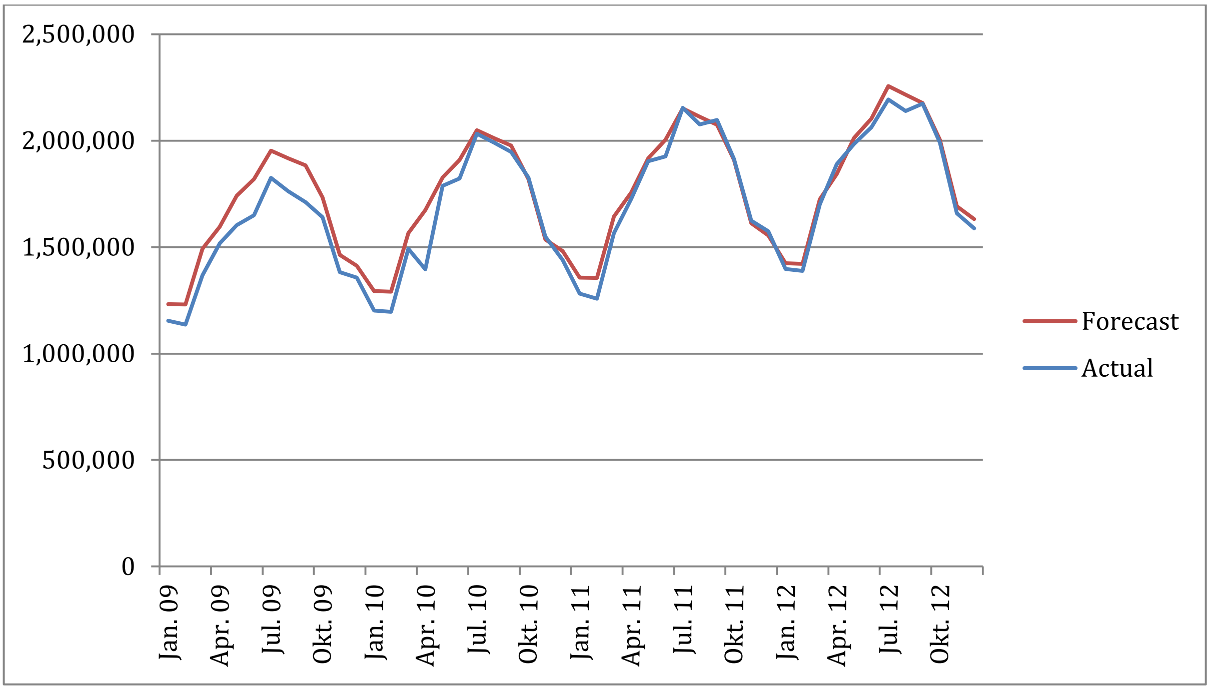
The graph of the five-year forecast can be seen in Figure 33 on the next page. Similar to the one-year prediction this one starts very accurately. However, it can be observed that after the month of June it starts to highly deviate from the actual values, although the used ARIMA (0,1,1)(0,1,1) model was considered by the ACF of the residuals and the chi-square test to be as appropriate as for the previous predictions. The reason for this big difference between the forecasted and the actual values is the global financial crisis. The model was calibrated with the data until 2008 and is predicting with the high trends of the years before. This example shows that the aviation business is not as secure as most people think and one has to consider such external factors. It also shows that the historical data in the model should be updated as often as possible and if one sees that the predictions are not as good as expected, than the model should be checked again for accuracy and maybe the parameters should be changed.

Figure 33: Out of sample test for the five-years forecast done by the ARIMA (0,1,1)(0,1,1) model



To illustrate how important it is to keep the model up-to-date with the newest figures available, a four-year forecast is also made. The graph can be seen in Figure 34 on the next page. Firstly one sees that the forecasted values for the years 2010, 2011 and 2012 are almost identical to the actual passenger numbers volumes. The fact that the prediction is that accurate is due to luck, but the important thing is that one is able to foresee the demands much better than in “the previous year”. The reason for this is that the model is updated with the historical data from the year 2008, where the crisis strikes and the passenger numbers started to drop after the month of May. By taking a look at those forecasted values one can easily see that the one-year forecast (the forecast for the year 2009) is not as accurate as the forecast shown in Figure 32. This is due to the fact that the passenger numbers continued to suffer even more than in 2008, something that couldn’t be expected by the model.

Figure 34: Out of sample test for the four-year forecast done by the ARIMA (0,1,1)(0,1,1) model



4 Conclusion

This work has discussed the construction, calibration and validation processes of two forecasting techniques that can be adopted by airports in order to predict their future passenger demands based only on passenger numbers historical data. It has also provide a description and explanation on how the models are to be used in practice by making a forecast for the Vienna International Airport with both of them.

Firstly an outline of the definition and the theoretical and mathematical fundaments of the rational passenger numbers planning technique are shown. The construction stages of the multi-period planned passenger numbers trajectory and the 90% enterprise passenger numbers uncertainty corridor, that accompanies the plan, were given to show how the fraction based enterprise passenger numbers model allow the formation of model-based passenger numbers forecast in a stochastic business environment. The construction processes are followed by a calibration of the model with the historical data of the Vienna International Airport. This provides a forecast for the airport and gives a better understanding of the correct use of the model. The adequacy of the model is than checked in an out of sample test to show that the model is appropriate and can be used for future passenger demands predictions. The last section of the first chapter showed how the Gaussian passenger numbers model is used for the process of subsequent planning to make sure that the planned demands are reached as accurately as possible.

The second part of the work begins with a detailed description and explanation of the theory behind the forecasting model ARIMA. It then provides a step-by-step guide to the methodology that should be adapted when using such model. This part then presents the actual fitting of the passengers' data and how it satisfies the forecasting expectations. This results in the choice of using a limited set of data and ignoring the initial years due to a number of external factors. The work then concludes that the chosen model is adequate and reliable for such set of variables and can be used by the business. However, it is worth mentioning that such predictions are largely based on historical data thus it is crucial to keep this data timely. Predictions should be made as often as possible in order to provide the most accurate and up to date predictions capturing the effects of external factors and unforeseen events.

5 References

Arens, T., Hettlich, F., Karpfinger, C., Kockelkorn, U., Lichtenegger, K. and Stachel, H. (2008) Mathematik, Heidelberg: Spektrum Akademischer Verlag

Benninga, S. and Wiener, Z. (1998) *Value-at-Risk (VaR)*, Mathematica in Education and Research Vol. 7 No. 4

Black, F. and Scholes, M. (1973). *The Pricing of Options and Corporate Liabilities*. Journal of Political Economy

Box, G.E.P. Jenkins, G.M. and Reinsel G.C. (2008) *Time Series Analysis: Forecasting and Control*, 4th edition, New Jersey: John Wiley & Sons, Inc.

Brockwell, P. and Davis, R. (2002) *Introduction to Time Series and Forecasting*, 2nd edition, New York: Springer Verlag

Cleff, T. (2008) *Deskriptive Statistik und modern Datenanalyse,* Wiesbaden: GWV Fachverlage GmbH

Cryer, J. and Chan, K. (2008) *Time Series Analysis*, 2nd edition, Iowa City: Springer Science+Business Media

Guharati, D. (2010) *Basic econometrics*, 5th edition, UK- McGraw-Hill

Hanke, J and Reitsch, A. (1998) *Business Forecasting*, 6th edition, Upper Saddle River: Prentice Hall.

Kelley, C.T. (1995) *Methods for Linear and Nonlinear Equations*, SIAM

Lucas, R. (1978). *Asset Prices in an Exchange Economy*, Econometrica

Markowitz, H. (1952). *Portfolio Selection*, Journal of Finance

Markowitz, H. (1959). *Portfolio Selection- Efficient Diversification of Investments*, New York: John Willey & Sons, Inc

Merton, R. (1973) *An Intertemporal Capital Asset Pricing Model*, Econometrica

Schwaiger, W. (2012) *Rational Sales Planning in the Probabilistic Management Framework, Institute of Management Sciences, Vienna University of Technology*

Schwaiger (2013) *Enterprise Risk Management- Foundamentals, Vienna University of Technology: Institute of Management Science*

Schwaiger, W., Bös A. and Kronfellner, B. (2012) *Risk-based Performance Management, Vienna University of Technology: Institute of Management Science*

Statistik Austria, (1964-2012) *Statistik der Zivilluftfahrt*

6 Appendix

1. *Normal distributions table*

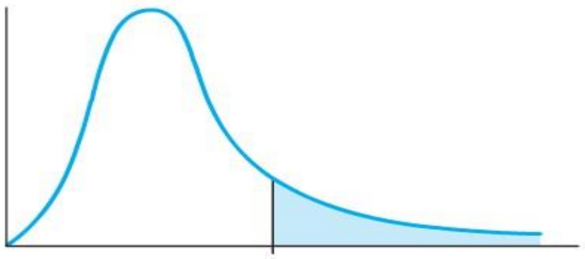
Table 18: Normal distributions table

z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	50,00%	50,40%	50,80%	51,20%	51,60%	51,99%	52,39%	52,79%	53,19%	53,59%
0,1	53,98%	54,38%	54,78%	55,17%	55,57%	55,96%	56,36%	56,75%	57,14%	57,53%
0,2	57,93%	58,32%	58,71%	59,10%	59,48%	59,87%	60,26%	60,64%	61,03%	61,41%
0,3	61,79%	62,17%	62,55%	62,93%	63,31%	63,68%	64,06%	64,43%	64,80%	65,17%
0,4	65,54%	65,91%	66,28%	66,64%	67,00%	67,36%	67,72%	68,08%	68,44%	68,79%
0,5	69,15%	69,50%	69,85%	70,19%	70,54%	70,88%	71,23%	71,57%	71,90%	72,24%
0,6	72,57%	72,91%	73,24%	73,57%	73,89%	74,22%	74,54%	74,86%	75,17%	75,49%
0,7	75,80%	76,11%	76,42%	76,73%	77,04%	77,34%	77,64%	77,94%	78,23%	78,52%
0,8	78,81%	79,10%	79,39%	79,67%	79,95%	80,23%	80,51%	80,78%	81,06%	81,33%
0,9	81,59%	81,86%	82,12%	82,38%	82,64%	82,89%	83,15%	83,40%	83,65%	83,89%
1,0	84,13%	84,38%	84,61%	84,85%	85,08%	85,31%	85,54%	85,77%	85,99%	86,21%
1,1	86,43%	86,65%	86,86%	87,08%	87,29%	87,49%	87,70%	87,90%	88,10%	88,30%
1,2	88,49%	88,69%	88,88%	89,07%	89,25%	89,44%	89,62%	89,80%	89,97%	90,15%
1,3	90,32%	90,49%	90,66%	90,82%	90,99%	91,15%	91,31%	91,47%	91,62%	91,77%
1,4	91,92%	92,07%	92,22%	92,36%	92,51%	92,65%	92,79%	92,92%	93,06%	93,19%
1,5	93,32%	93,45%	93,57%	93,70%	93,82%	93,94%	94,06%	94,18%	94,29%	94,41%
1,6	94,52%	94,63%	94,74%	94,84%	94,95%	95,05%	95,15%	95,25%	95,35%	95,45%
1,7	95,54%	95,64%	95,73%	95,82%	95,91%	95,99%	96,08%	96,16%	96,25%	96,33%
1,8	96,41%	96,49%	96,56%	96,64%	96,71%	96,78%	96,86%	96,93%	96,99%	97,06%
1,9	97,13%	97,19%	97,26%	97,32%	97,38%	97,44%	97,50%	97,56%	97,61%	97,67%
2,0	97,72%	97,78%	97,83%	97,88%	97,93%	97,98%	98,03%	98,08%	98,12%	98,17%
2,1	98,21%	98,26%	98,30%	98,34%	98,38%	98,42%	98,46%	98,50%	98,54%	98,57%
2,2	98,61%	98,64%	98,68%	98,71%	98,75%	98,78%	98,81%	98,84%	98,87%	98,90%
2,3	98,93%	98,96%	98,98%	99,01%	99,04%	99,06%	99,09%	99,11%	99,13%	99,16%
2,4	99,18%	99,20%	99,22%	99,25%	99,27%	99,29%	99,31%	99,32%	99,34%	99,36%
2,5	99,38%	99,40%	99,41%	99,43%	99,45%	99,46%	99,48%	99,49%	99,51%	99,52%
2,6	99,53%	99,55%	99,56%	99,57%	99,59%	99,60%	99,61%	99,62%	99,63%	99,64%
2,7	99,65%	99,66%	99,67%	99,68%	99,69%	99,70%	99,71%	99,72%	99,73%	99,74%
2,8	99,74%	99,75%	99,76%	99,77%	99,77%	99,78%	99,79%	99,79%	99,80%	99,81%
2,9	99,81%	99,82%	99,82%	99,83%	99,84%	99,84%	99,85%	99,85%	99,86%	99,86%
3,0	99,87%	99,87%	99,87%	99,88%	99,88%	99,89%	99,89%	99,89%	99,90%	99,90%

2. *Critical Values of Chi-Square*

Figure 35: Critical Values of Chi-Square

*The table entries are critical values of χ^2.



df	Proportion in Critical Region 0.10	0.05	0.025	0.01	0.005
1	2.71	3.84	5.02	6.63	7.88
2	4.61	5.99	7.38	9.21	10.60
3	6.25	7.81	9.35	11.34	12.84
4	7.78	9.49	11.14	13.28	14.86
5	9.24	11.07	12.83	15.09	16.75
6	10.64	12.59	14.45	16.81	18.55
7	12.02	14.07	16.01	18.48	20.28
8	13.36	15.51	17.53	20.09	21.96
9	14.68	16.92	19.02	21.67	23.59
10	15.99	18.31	20.48	23.21	25.19
11	17.28	19.68	21.92	24.72	26.76
12	18.55	21.03	23.34	26.22	28.30
18	25.99	28.87	31.53	34.81	37.16
19	27.20	30.14	32.85	36.19	38.58
20	28.41	31.41	34.17	37.57	40.00
21	29.62	32.67	35.48	38.93	41.40
22	30.81	33.92	36.78	40.29	42.80
23	32.01	35.17	38.08	41.64	44.18
24	33.20	36.42	39.36	42.98	45.56
25	34.38	37.65	40.65	44.31	46.93
26	35.56	38.89	41.92	45.64	48.29
27	36.74	40.11	43.19	46.96	49.64
28	37.92	41.34	44.46	48.28	50.99
29	39.09	42.56	45.72	49.59	52.34
30	40.26	43.77	46.98	50.89	53.67
40	51.81	55.76	59.34	63.69	66.77
50	63.17	67.50	71.42	76.15	79.49
60	74.40	79.08	83.30	88.38	91.95
70	85.53	90.53	95.02	100.42	104.22
80	96.58	101.88	106.63	112.33	116.32
90	107.56	113.14	118.14	124.12	128.30
100	118.50	124.34	129.56	135.81	140.17

2. *Tables for the Subsequent Forecasting of January and February 2012*

Table 19: Subsequent Forecasting January and February 2012

After Month 1 - 2012														
Model	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12	Total	Notes
$x^{Plan}_{P(PF)}$													22.448.524	Planned Annual N
$x^{A}_{P(PF),M}$	1.397.692													Actual Monthly PN
$AA_{P(PF)}(s_t)$	1.397.692													Accumulated Actual PN
$\mu_{a(PF),M}(s_t)$		6,07%	7,47%	8,25%	8,97%	9,15%	10,18%	9,87%	9,72%	9,09%	7,70%	7,39%		Planned Monthly Fraction
$\mu_{a(PF),T}(s_t)$													93,85%	Cond. Remaining Fraction
$\mu_{P(PF),T}(s_t)$													21.068.942	Cond. Remaining Expectations
$\mu_{P(PF)}(s_t)$													22.466.634	Cond. PN Forecast
$PFD_{P(PF)}(s_t)$													18.110	Cond. Plan/Forecast Deviation
$\sigma_{a(PF),M}(s_t)$		0,879%	0,828%	0,976%	0,412%	0,442%	0,787%	0,802%	0,960%	0,777%	0,636%	0,703%		Planned Mont. Fraction Vola
$\sigma_{a(PF),T}(s_t)$													2,54%	Cond. Remaining Fraction Vola
$\sigma_{P(PF),T}(s_t)$													570.592	Cond. Remaining PN Vola
$PFD^{RA}_{P}(s_t)$													0,031739333	Risk Adjusted PFD (z-value)
$p(z_{PNs})(s_t)$													51,27%	p-Value of z-Value

After Month 2 - 2012														
Model	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12	Total	Notes
$x^{Plan}_{P(PF)}$													22.448.524	Planned Annual N
$x^{A}_{P(PF),M}$	1.397.692	1.387.505												Actual Monthly PN
$AA_{P(PF)}(s_t)$	1.397.692	2.785.197												Accumulated Actual PN
$\mu_{a(PF),M}(s_t)$			7,47%	8,25%	8,97%	9,15%	10,18%	9,87%	9,72%	9,09%	7,70%	7,39%		Planned Monthly Fraction
$\mu_{a(PF),T}(s_t)$													87,78%	Cond. Remaining Fraction
$\mu_{P(PF),T}(s_t)$													19.706.186	Cond. Remaining Expectations
$\mu_{P(PF)}(s_t)$													22.491.383	Cond. PN Forecast
$PFD_{P(PF)}(s_t)$													42.859	Cond. Plan/Forecast Deviation
$\sigma_{a(PF),M}(s_t)$			0,828%	0,976%	0,412%	0,442%	0,787%	0,802%	0,960%	0,777%	0,636%	0,703%		Planned Mont. Fraction Vola
$\sigma_{a(PF),T}(s_t)$													2,39%	Cond. Remaining Fraction Vola
$\sigma_{P(PF),M}(s_t)$													535.428	Cond. Remaining PN Vola
$PFD^{RA}_{P}(s_t)$													0,080046606	Risk Adjusted PFD (z-value)
$p(z_{PNs})(s_t)$													53,19%	p-Value of z-Value

7 List of figures

8 List of Tables

9 Table of abbreviations

ACF	Autocorrelation function
AR	Autoregressive
ARIMA	Autoregressive integrated moving-average
Cov	Covariance
Etc.	Et cetera
EU	European Union
MA	Moving-average
NS	Non-Schengen
p.	Page
PACF	Partial Autocorrelation function
PFD	Plan Forecast Deviation
PN	Passenger Numbers
pp.	Page
ROW	Rest of the World
S	Schengen
VIE	Vienna International Airport
Vola	Volatility