Computer Modelling and New Technologies, 2007, Vol.11, No.1, 62-72
Transport and Telecommunication Institute, Lomonosova 1, Riga, LV-1019, Latvia

USE OF THE GENERALIZED LINEAR MODEL IN FORECASTING THE AIR PASSENGERS' CONVEYANCES FROM EU COUNTRIES

Catherine Zhukovskaya

Riga Technical University, Faculty of Transport and Mechanical Engineering
Kalku 1, LV-1658, Riga, Latvia
E-mail: kat_zuk@hotmail.com

Some regression models to forecast the air passengers' conveyances from EU countries are considered. Two different approaches for the above-mentioned task of forecasting are shown. The first one is the classical method of *linear regression* and the second one is its *generalized* approach. The considered regression models contain many explanatory factors and their combinations. The advantage of using the *generalized linear model* (GLM) in comparison with the classical *linear regression model* is shown.

Keywords: *air passengers' conveyances, forecasting, generalized linear model*

1. Introduction

Most the literature which is devoted to forecasting of transport flows contains only simple forecasting models on the base of the time series methods or linear regression methods with a small number of explanatory variables. Two approaches for the forecasting of air passengers' conveyances from EU countries are considered in this article: the classical method of *linear regression* and its *generalized* approach. The difference of linear regression models considered in this article comparing with the models presented in other papers [6] (autoregression integrated moving average models) and [3, 4, 8] (multiple regression models) consists in using the greater number of the explanatory factors and their combinations. Some models on the base of GLM are considered in the article as well. The aim of this article is to illustrate the advantage of using the GLM comparing with the linear regression models. The verification of the models and the evaluation of the unknown parameters are included in the research as well.

This article has the following structure. The second section contains the description of the informative base of the mentioned investigation. The used models for analyzing and forecasting of air passengers' conveyances are considered in the third section. The elaboration of linear regression models and generalized linear models are presented in the fourth and fifth sections. In the fifth section the advantage of using GLM in comparison with the classical linear regression is shown.

2. Informative Base

In this article the number of carried air passengers was our index of interest and we intend to forecast their volumes. We use the following factors influencing the volumes of air passengers' conveyances:

t_1 total population of the country (TP), millions of inhabitants;
t_2 area of the country (AREA), thousands of km^2;
t_3 density of the country population (PD), number of inhabitants per km^2;
t_4 monthly labour costs (MLC), thousands of euro;
t_5 gross domestic product (GDP) "per capita" in Purchasing Power Standards (PPS) (GDP_PPS);
t_6 gross domestic product (GDP), billions of euro;
t_7 comparative price level (CPL);
t_8 inflation rate (IR);
t_9 unemployment rate (UR);
t_{10} labour productivity per hour worked (LPHW).

The time interval of consideration was the period from 1996 to 2005. We consider the air passengers' conveyances from EU countries. By the moment of data collection there were 25 countries in the EU, such as Belgium, Czech Republic, Denmark, Germany, Estonia, Greece, Spain, France, Ireland, Italy, Cyprus, Latvia, Lithuania, Luxembourg, Hungary, Malta, Netherlands, Austria, Poland, Portugal, Slovenia, Slovakia, Finland, Sweden and the United Kingdom. All data for this investigation have been received from the electronic database "The Statistical Office of the European Communities" (EUROSTAT) [9].

Some of the considered above factors are to be commented:

a) GDP per capita (Latin: *for each head*) in PPS is the value of all final goods and services produced within a nation in a given year divided by the average population for the same year. This volume index of GDP is expressed in relation to the European Union (EU25 = 100).

b) Comparative price level is an index that used for cross-country comparison of price levels. If it is higher/lower than 100 (EU25 = 100), the country concerned is relatively expensive/cheap as compared with the EU average.

For each considered country and for each year we have the volumes of all ten basic factors mentioned above and the volumes of raw conveyances of the air passengers' carried. But during the data gathering we have collided with shortage of data on many countries, especially concerning the new members of EU; therefore the final number of the observation was 161.

The data for the period from 1996 to 2004 have been used for the estimation and forecasting, i.e. for finding of coefficients of the regressional models (140 observations). The data of the 2005 (21 observations) have been used to check out the quality of forecasting, the so-called the cross-validation (CV). Detailed description of CV approach is considered by Diana Santalova in the proceeding article [7].

3. The Used Models for Analyzing and Forecasting of the Air Passengers' Conveyances

The air passengers' conveyances from EU countries were the main *object* of the consideration in our investigation. The data about concrete country for the concrete year were taken as the *observation*. All the considered models were the *group models* [1]. It means that we have the identical regressional model for the various similar objects.

In our research the *linear regression models* and the *generalized regression models* have been used. In the simplest case *the linear regression model* can be expressed in the following form [5]

$$E\left(Y^{(k)}(x)\right) = \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}, \quad (1)$$

where $Y^{(k)}$ is a dependent variable for the k-th considered model (regressand), $\mathbf{x} = (x_1, x_2, ..., x_d)^T$ is d-dimensional vector of regressors or explanatory variables, $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, ..., \beta_d)^T$ is a coefficient vector that has to be estimated from observations for $Y^{(k)}$ and $\mathbf{x}$.

The great number of linear regressional models [3, 4, 8] offered in the literature contains small number of the explanatory variables. But just increasing their number does not lead to improving considerably the quality of the regressional models. So the *generalized linear model* can be applied [5]:

$$E\left(Y^{(k)}(x)\right) = G\left\{\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}\right\}, \quad (2)$$

where $G(\circ)$ is the *known link function* of one dimensional variable.

Firstly, we consider the linear regressional models. After that the generalized linear models are in focus of our research.

4. Elaboration of Linear Models

The big number of various models which differed with structure of factors and their combinations has been constructed and investigated. During the process of the models' construction the received results have been constantly analyzed and the new complementary factors have been added to them. All the considered models in this investigation are the group models.

As the basic criteria to choose the best model, the following ones were selected: *the multiple coefficient of determination* (R^2), *Fisher's criterion* (F), *the sum of the squares of the residuals* (SSRes) and *the sum of the squares of residuals for the cross-validation* (CV SSRes). In additional to these criteria the other ones have been considered as well, in particular the forms of bands of residuals have been analyzed. Let us note that for the checking of the statistical hypotheses we always used the statistical significance level $\alpha = 0.05$.

In the models 1-3 as the regressand $Y^{(i)} = y$ (where $i = 1, 2, 3$) the number of raw air passengers' carried was taken.

As the regressors in Model 1 all the considered above variables without their modification such as: $x_1 = t_1$, $x_2 = t_2$, $x_3 = t_3$, $x_4 = t_4$, $x_5 = t_5$, $x_6 = t_6$, $x_7 = t_7$, $x_8 = t_8$, $x_9 = t_9$ and $x_{10} = t_{10}$ were chosen.

Model 1 gives the following estimate for $E(Y)$:

$$\hat{E}\left(Y^{(1)}(x)\right) = 14 - 0.77x_1 + 0.16x_2 + 185.8x_3 - 2.44x_4 + 0.53x_5 + 0.07x_6 + 0.05x_7 + 0.32x_8 - \\ -1.2x_9 - 1.03x_{10}.$$

Now we are going to consider some criteria for the used model. The value of the coefficient of determination for this model $R^2 = 0.831$ is high enough. The value of the Fisher criterion $F = 63.49$ is considerably higher than its critical value $F_k = 1.905$. This value was calculated with the significance level $\alpha = 0.05$ and with the degrees of freedom $df1 = m = 10$ and $df2 = (n - m - 1) = 129$, where m is a number of predicted values and n is a number of observations [2]. The critical level of model significance (*p-level*) p = 0.000000, so this model is adequate.

For each factor of model 1 table 1 contains factor estimates (b) and the results of the check of their significance: the calculated values of the Student *t*-criterion (t) and p-level. Some factors for this model are nonsignificant. The critical value of the 2-tailed Student criterion $t_k = 1.979$ which was calculated with the significance level $\alpha = 0.05$ and with the degrees of freedom $df = (n - m - 1) = 129$ [2]. The significant explanatory variables are the variables x_2, x_3, x_6 and x_{10}, so, the greatest influence on the air passengers' conveyances is provided by the aria of the country, the density of the country population, the value of the gross domestic product and the comparative price level. The positive and the negative signs for all regressors in this model correspond to physical sense of regressors. Such statistical procedure was used for all linear regression models considered below.

TABLE 1. The estimates of the coefficients and their significance level for Model 1

Variable	Factor	b	t(129)	p-level
	Intercept	14.00	0.84	0.405
x_1	TP	-0.77	-1.56	0.121
x_2	AREA	0.16	5.60	0.000
x_3	PD	185.80	4.67	0.000
x_4	MLC	-2.44	-0.44	0.660
x_5	GDP_PPS	0.53	1.68	0.096
x_6	GDP	0.07	3.81	0.000
x_7	CPL	0.05	0.37	0.710
x_8	IR	0.32	0.29	0.771
x_9	UR	-1.20	-1.59	0.114
x_{10}	LPHW	-1.03	-3.75	0.000

The analysis of the form of the band of residuals for Model 1 has shown the necessity of adding into the regression model the new explanatory factor t_{11} (ON). This factor takes 2 meanings: "0" if the considered country is the old member of EU, and "1" if the considered country is the new one. Additionally we remove some nonsignificant factors from Model 1.

Therefore the regressors in Model 2 are the following: $x_1 = t_2$, $x_2 = t_3$, $x_3 = t_6$ $x_4 = t_{10}$ and $x_5 = t_{11}$.

Model 2 gives the following estimate for $E(Y)$:

$$\hat{E}\left(Y^{(2)}(x)\right) = 13.56 + 0.09x_1 + 134.01x_2 + 0.05x_3 - 0.68x_4 + 29.36x_5.$$

The obtained results for Model 2 are shown in the Table 2. We see that our modification allows improving some characteristics of regressional Model 1. In this model $R^2 = 0.829$ but despite the decrease of it we see that the value of $F = 129.85$ has considerably increased comparing with the previous model.

TABLE 2. The estimates of the coefficients and their significance level for Model 2

Variable	Factor	b	t(134)	p-level
	Intercept	13.56	2.45	0.016
x_1	AREA	0.09	4.45	0.000
x_2	PD	134.01	4.32	0.000
x_3	GDP	0.05	10.34	0.000
x_4	LPHW	-0.68	-5.12	0.000
x_5	ON	29.36	4.21	0.000

The next step for improving the characteristics of the regressional models consists in adding to regressional models the modified basic factors and their different combinations, such as: $\sqrt{t_1}$, t_1^2, $\sqrt{t_2}$, t_2/t_1, $\sqrt{t_2}/t_1$, t_6/t_1, $t_6/(t_1 \cdot t_2)$, $t_6/(t_1 \cdot \sqrt{t_2})$. We begin with Model 3, where: $x_1 = t_3$, $x_2 = t_6$, $x_3 = t_{10}$, $x_4 = t_1^2$ and $x_5 = \sqrt{t_2}$.

This model gives the following estimate for $E(Y)$:

$$\hat{E}\left(Y^{(3)}(x)\right) = -6.34 + 113.26x_1 + 0.14x_2 - 0.52x_3 - 0.03x_4 + 3.03x_5.$$

The analysis of the obtained results for Model 3 (Table 3) also shows the rightful appliance of this approach because it allows us to improve considerably the characteristics of the regressional model. Moreover, the comparison of the received results with the results which have been obtained for the models considered above shows that their input allows to improve two characteristics of the regression model at the same time: $R^2 = 0.867$ and $F = 174.078$.

TABLE 3. The estimates of the coefficients and their significance level for Model 3

Variable	Factor	b	t(134)	p-level
	Intercept	-6.34	-1.05	0.296
x_1	PD	113.26	4.00	0.000
x_2	GDP	0.14	10.66	0.000
x_3	LPHW	-0.52	-5.80	0.000
x_4	sq(TP)	-0.03	-7.56	0.000
x_5	sqrt(AREA)	3.03	5.74	0.000

The observed and predicted values of the air passengers' conveyences in Country-Year order for Model 3 are shown on Figure 1, the results of the cross-validation for this model are shown on Figure 2. The "Country-Year order" for all values which is shown on Figure 1 means that, firstly, they are sorted by the country name and for each country they are sorted by year.

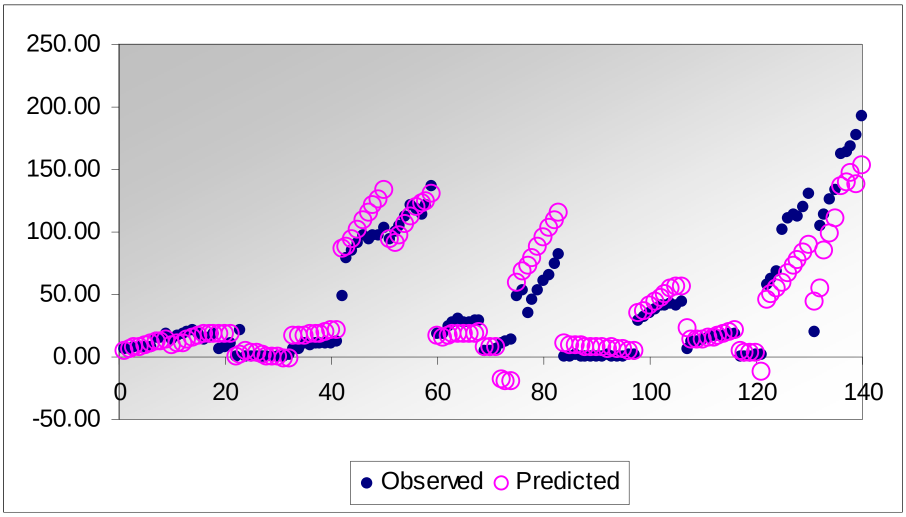

Figure 1. Plot of the observed and predicted values for Model 3

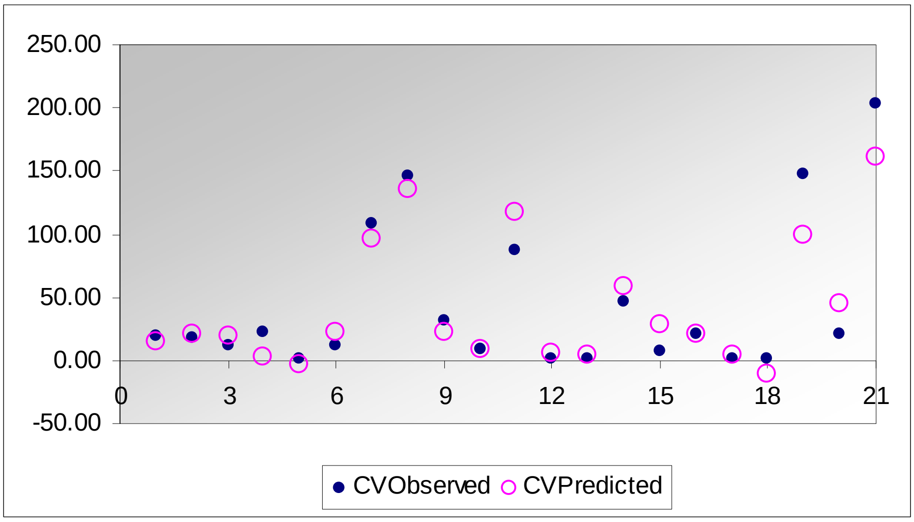

Figure 2. Plot of the observed and predicted values for the CV for Model 3

But in Figures 1 and 2 we can see considerable inconvenience of the 3 Model, which consists in the fact that some predicted values for this model, lies in the negative area. Models 1 and 2 have the same disadvantage. Therefore as the next step for the improving of the regressional model was the transfer to the new forecasting variables.

So in the Models 4-5 as the regressand we considered the ratio between the total number of air passenger carried and the number of inhabitants of the country $Y^{(i)} = y/t_1$, $i = 4, 5$.

As the regressors in Model 4 we used the following variables: $x_1 = t_2$, $x_2 = t_3$, $x_3 = t_4$, $x_4 = t_6$, $x_5 = t_{11}$, $x_6 = \sqrt{t_1}$, $x_7 = \sqrt{t_2}$, $x_8 = t_2/t_1$, $x_8 = \sqrt{t_2}/t_1$ and $x_9 = t_6/t_1$.

Model 4 gives the following estimate for $E(Y)$:

$$\hat{E}\left(Y^{(4)}(x)\right) = 0.56 + 2.33x_1 - 1.04x_2 - 0.02x_3 + 0.001x_4 + 1.76x_5 - 0.0004x_6 + 0.04x_7 + 0.17x_8.$$

The obtained results for Model 4 are shown in the Table 4 (R^2= 0.760, F = 45.81).

TABLE 4. The estimates of the coefficients and their significance level for Model 4

Variable	Factor	b	t(131)	p-level
	Intercept	-5.67	-6.25	0.000
x_1	AREA	-0.02	-6.73	0.000
x_2	PD	10.37	6.19	0.000
x_3	MLC	-0.73	-4.19	0.000
x_4	ON	0.83	8.30	0.000
x_5	sqrt(TP)	-1.02	-7.32	0.000
x_6	sqrt(AREA)	1.06	7.10	0.000
x_7	AREA/TP	-0.12	-6.98	0.000
x_8	sqrt(AREA)/TP	0.94	5.84	0.000
Variable	Factor	b	t(131)	p-level

The analysis of the obtained results doesn't show the improvement of the characteristics of this model. Therefore we decide to enter one more variable t_{12} (HL), which expresses the relative value of the conveyances. It takes 2 meanings: 0 if the value of y/t_1 for the considered country is small (less than 2) and is equal 1 if this value is larger than 2

As the regressors in Model 5 we used the following variables: $x_1 = t_4$, $x_2 = t_5$, $x_3 = t_8$, $x_4 = t_9$, $x_5 = t_{10}$, $x_6 = t_{11}$, $x_7 = t_{12}$ and $x_8 = t_6/t_1$.

Model 5 gives the following estimate for $E(Y)$:

$$\hat{E}\left(Y^{(5)}(x)\right) = 0.99 - 0.46x_1 - 0.02x_2 - 0.02x_3 - 0.02x_4 + 0.01x_5 + 1.27x_6 + 1.15x_7 + 0.07x_8 .$$

The obtained results for Model 5 are shown in the Table 5 (R^2 = 0.864, F = 104.174).

TABLE 5. The estimates of the coefficients and their significance level for Model 5

Variable	Factor	b	t(131)	p-level
	Intercept	0.99	3.93	0.000
x_1	MLC	-0.46	-3.41	0.001
x_2	GDP_PPS	-0.02	-3.81	0.000
x_3	IR	-0.02	-1.33	0.187
x_4	UR	-0.02	-1.90	0.056
x_5	LPHW	0.01	3.72	0.000
x_6	ON	1.27	9.21	0.000
x_7	HL	1.15	15.30	0.000
x_8	GDP/TP	0.07	3.41	0.001

The data for all considered models and for the four mentioned above criteria have been brought in the Table 6. For each model and for each criterion the rank R_i (where i = 1, 2, 3, 4) has been calculated. Here, the rank of a model is its i-th criterion position number among all values of this criterion. The sum of the ranks (Sum R) for all four criteria and the total rank (Total R) has allowed us to define the best model. In order to compare the results obtained for the Models 4-5 with the previous ones (Models 1-3), for the Models 4-5 the recalculated data for SSRes and CV SSRes have been used. These data were multiplied by the value of the country population. So according to the sum of ranks for all considered models and taking into account the inconvenience of the first three models we can conclude that the best model is Model 5.

TABLE 6. Pivot results for the first three models

Model	R^2	R_1	F	R_2	SSRes	R_3	CV SSRes	R_4	Sum R	Total R
Model 1	0.8311	3	63.49	4	52651.33	5	17232.75	5	17	5
Model 2	0.8289	4	129.85	2	53343.53	5	16458.41	4	15	3
Model 3	0.8666	1	174.1	1	41598.60	2	7417.482	1	5	1
Model 4	0.7603	5	45.81	5	35064.04	3	8596.43	3	16	4
Model 5	*0.8642*	*2*	*104.2*	*3*	*12774.59*	*1*	*7717.23*	*2*	*8*	*2*

Figure 3 shows the recalculated observed and predicted values for the air passengers' conveyances for Model 5 in the order Country-Year. The result of the cross-validation for this model is shown in Figure 4.

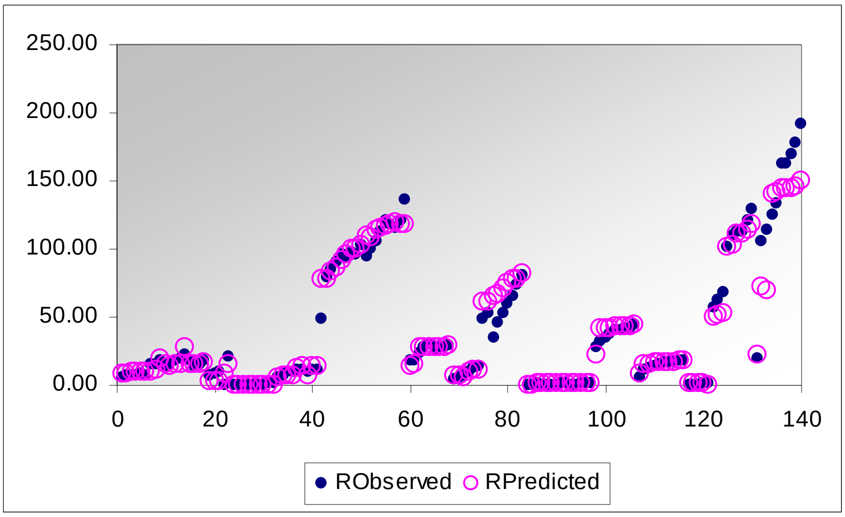

Figure 3. Plot of the recalculated observed and predicted values for Model 5

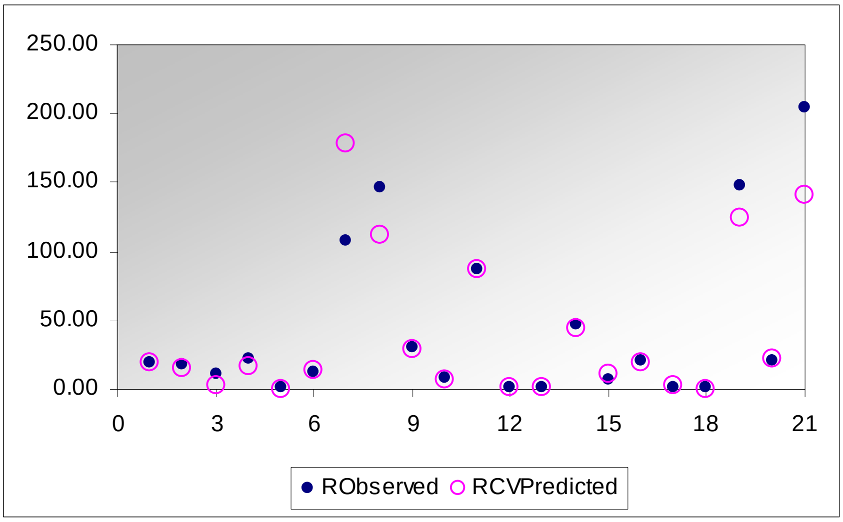

Figure 4. Plot of the recalculated observed and predicted values for the CV for Model 5

5. Elaboration of Generalized Linear Models

Tempting to improve the characteristics of the received linear regression models (1) the generalized linear model (2) has been used.

For the further investigation the best linear regression model (Model 5) has been chosen. Two different GLM are considered. In both of them the value of the regressand $Y^{(GLM)} = Y^{(5)}/t_1$ and the collection of the regressors are the same as for Model 5.

The first GLM (GLM1) is *the modification of the logit model* [5] which can be written in the following form:

$$E\left(Y^{(GLM1)}(x_i)\right) = h_i \frac{\exp\left(\sum_j \beta_j x_{i,j}\right)}{1 + \exp\left(\sum_j \beta_j x_{i,j}\right)}, \tag{3}$$

where h_i is the total population number, x_i is vector-columns of the independent variables, i is the observation number, $i = 1, 2, ..., n$.

The second of the investigated GLM (GLM2) has the following form:

$$E\left(Y^{(GLM2)}(x_i)\right) = h_i \frac{1}{a + \exp\left(\sum_j \beta_j x_{i,j}\right)}, \tag{4}$$

where a is the additional parameter (constant).

The unknown parameter vector $\boldsymbol{\beta} = (\beta_1, \beta_2, ..., \beta_d)^T$ for both GLM is estimated by the use of the least squares (LS) criterion:

$$R_0(\beta) = \sum_{i=1}^{n} \left(Y_i - \tilde{Y}_i\right)^2 \to \min_{\beta}, \tag{5}$$

where Y_i and $\tilde{Y}_i$ are observed and calculated values of Y.

Note, that *linearization* of the logistics models is the traditional way for the estimation of their unknown parameters. Let us show that it gives bad results.

After GLM linearization their liniarizated forms (LM) LM1 for the model (3) and LM2 for the model (4) correspondingly were obtained:

for the model (3) the LM1

$$\ln \frac{Y^*}{1-Y^*} = \sum_j \beta_j x_{i,j} , \tag{6}$$

and for the model (4) the LM2

$$\ln\left(\frac{1}{Y^*} - a \right) = \sum_j \beta_j x_{i,j} , \tag{7}$$

where $Y^* = Y/h$.

Here the dependent variables are situated in the left side of the formulas, and their parameters' β estimation is not difficult to do.

The models LM1 and LM2 gives the following estimate for $E(Y)$:

$$\hat{E}\left(Y^{(LM1)}(x)\right) = h \frac{e^{-13.78+0.001x_1-6.68x_2-0.02x_3+0.7x_4+48.8x_5-0.44x_6+0.29x_7+7.81x_8-0.64x_9}}{1+e^{-13.78+0.001x_1-6.68x_2-0.02x_3+0.7x_4+48.8x_5-0.44x_6+0.29x_7+7.81x_8-0.64x_9}} ,$$

$$\hat{E}\left(Y^{(LM2)}(x)\right) = h \frac{1}{0.3+e^{11.65+1.63x_1-1.7x_2+0.04x_3-0.81x_4-17.96x_5-1.67x_6+0.2x_7+0.41x_8-0.11x_9}} .$$

The values of SSRes and CV SSRes for the Models LM1 and LM2 comparing with Model 5 are calculated and shown in the Table 7.

TABLE 7. The value of SSRes and CV SSRes for the Models 5, LM1 and LM2

	SSRes			**CV SSRes**		
	Model 5	LM1	LM2	Model 5	LM1	LM2
R_0/n	12 775	27 447	21 834	7 717	676 576	229 554

We can see that linearization gives bad results. Making attempts to improve the obtained results a two-stage estimation procedure is developed. The first stage corresponds to the considered linearization. As the second step we use the procedure of *calibration* when we precise the gotten estimates by using the well-known gradient method.

The gradients with the respect to the unknown parameter vector $\boldsymbol{\beta}$ for the GLM1 and GLM2 can be written in the following forms:
for the model (3)

$$\nabla R(\beta) = -2\sum_{i=1}^{n-1} \left(Y_i - h_i \frac{\exp\left(\sum_j \beta_j x_{i,j}\right)}{1+\exp\left(\sum_j \beta_j x_{i,j}\right)} \right) \cdot h_i \cdot \frac{\exp\left(\sum_j \beta_j x_{i,j}\right)}{\left(1+\exp\left(\sum_j \beta_j x_{i,j}\right)\right)^2} \cdot x_i , \tag{8}$$

and for the model (4)

$$\nabla R(\beta) = 2\sum_{i=1}^{n-1} \left(Y_i - h_i \frac{1}{a+\exp\left(\sum_j \beta_j x_{i,j}\right)} \right) \cdot h_i \cdot \frac{\exp\left(\sum_j \beta_j x_{i,j}\right)}{\left(a+\exp\left(\sum_j \beta_j x_{i,j}\right)\right)^2} \cdot x_i . \tag{9}$$

For the GLM2 we found the optimum value of R_0 not only from the values β but from the parameter a also. Firstly, we fix the value of a and search the optimum values of β according to (5). After that we fix the founded values of β and search the optimum values of R_0 by changing the value of parameter a. This procedure has been repeated many times until R_0 reaches its minimum.

The GLM1 and GLM2 have the following estimates for $E(Y)$:

$$\hat{E}\left(Y^{(GLM1)}(x)\right) = h\frac{e^{-7.05-1.05x_1+1.22x_2-0.02x_3+0.76x_4+5.77x_5+1.26x_6-0.11x_7-0.68x_8+0.15x_9}}{1+e^{-7.05-1.05x_1+1.22x_2-0.02x_3+0.76x_4+5.77x_5+1.26x_6-0.11x_7-0.68x_8+0.15x_9}},$$

$$\hat{E}\left(Y^{(GLM2)}(x)\right) = h\frac{1}{6.3+e^{7.26+1.09x_1-0.78x_2+0.02x_3-0.82x_4-7.81x_5-1.12x_6+0.1x_7+0.13x_8-0.06x_9}}.$$

Figure 5 shows the observed values of air passengers' conveyances and predicted values obtained by using of the generalized linear regression models GLM1 and GLM2 in the order to Country-Year. The results of the cross-validation for the models GLM1 and GLM2 are shown in Figure 6.

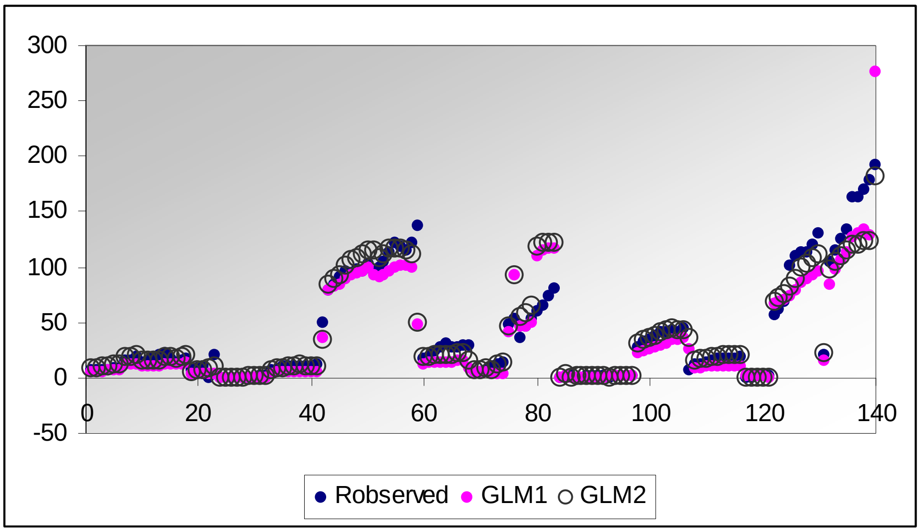

Figure 5. Plot of the observed and predicted values for the GLM1 and GLM2

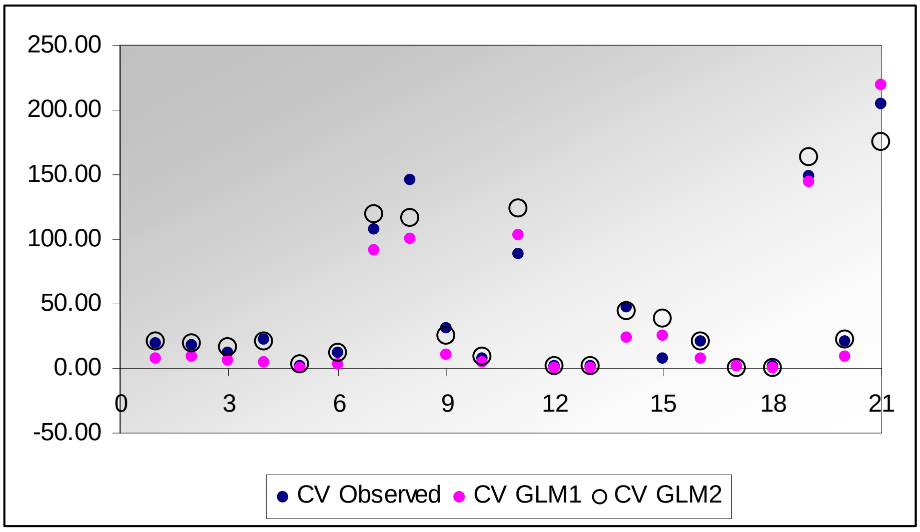

Figure 6. Plot of the observed and predicted values for the CV for the GLM1 and GLM2

The SSRes for the CV for the Models 5, GLM1 and GLM2 are shown in the Table 8.

TABLE 8. The value of SSRes for the CV for the Models 5, GLM1 and GLM2

	CV SSRes		
	Model 5	GLM1	GLM2
R_0/n	7 717	7 171	5 185

The comparison of the data from Tables 7 and 8 allows stating the following: generalized linear models give better results for the forecasting of air passengers' conveyances in comparison with traditional linear regression models; simple linearization gives considerably worse results for the forecasting and needs in its optimization. For this purpose the two-stage estimation procedure which is shown in this article can be used.

Besides this for the GLM2 the dependence of values SSRes and CV SSRes on the value of parameter α is investigated. The obtained results are shown in Figure 7. The optimal value for analysis of SSRes is obtained, then $\alpha = 2$, but the best result for the analysis of CV SSRes is obtained, then $\alpha = 6$.

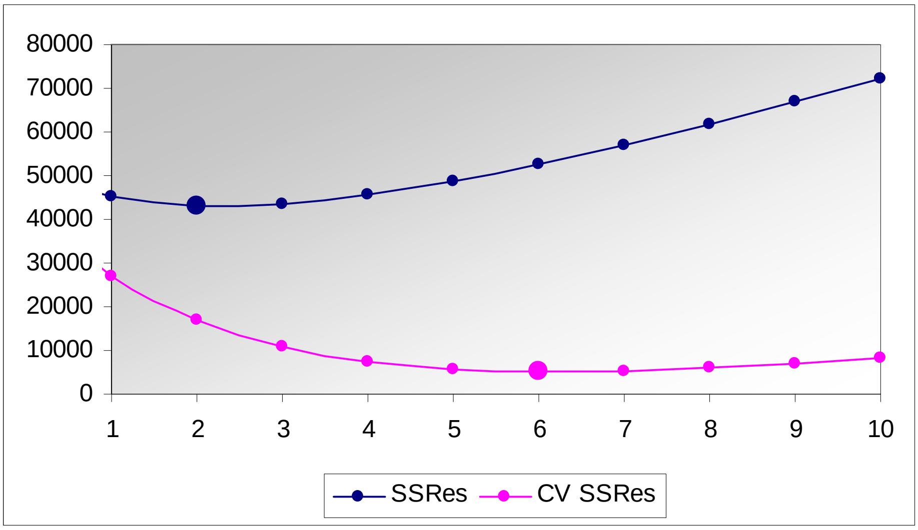

Figure 7. The values of SSRes and CV SSRes as a function of parameter α for GLM 2

Conclusion

The linear and generalized linear regressional models for the forecasting of air passengers'' conveyances from EU countries are considered. These models contain a big number of explanatory factors and their combinations. For the estimation of the unknown parameters of the linear regressional models we use the standard procedures. For the estimation of unknown parameters of GLM the special two-stage procedure has been elaborated. The cross-validation approach has been taken as the main procedure for the check out the adequacy of all considered models and choosing the best model for the forecasting. The advantage of GLM application has been shown.

References

1. Andronov, A. M. etc. *Forecasting of air passengers' conveyances on the transport.* Moscow: Transport, 1983. (In Russian)
2. Andronov, A. M, Kopytov, E. A and Gringlaz, L. J. *The probability theory and the mathematical statistics.* StPetersburg: Piter, 2004. (In Russian)
3. Butkevicius, J., Mazura, M., Ivankovas, V., Mazura, S. Analysis and forecast of the dynamics of passenger transportation by public land transport, *Transport*, Vol. XIX, No 1, 2004, pp. 3-8.
4. Butkevicius, J., Vyskupaitis, A. Development of passenger transportation by Lithuanian sea transport. In: *Proceedings of International Conference RelStat'04, Transport and Telecommunication,* Vol. 6, No 2, 2005.

5. Hardle, W., Muller, M., Sperlich, S., Werwatz, A. *Nonparametric and Semiparametric Models*. Berlin: Springer, 2004.
6. Hunt, U. Forecasting of railway freight volume: approach of Estonian railway to arise efficiency, *Transport*, Vol. XXVIII, No 6, 2003, pp.255-258.
7. Santalova, D. Forecasting of Rail Freight Conveyances in EU Countries on the Base of the Single Index Model. (In printing)
8. Sliupas, T. Annual average daily traffic forecasting using different techniques, *Transport*, Vol. XXI, No 1, 2006, pp.38-43.
9. *EUROSTAT YEARBOOK 2005. The statistical guide to Europe. Data 1993-2004. EU, EuroSTAT, 2005. URL –* http://epp.eurostat.ec.europa.eu