Martin Pavlik, PhD.

THE USAGE OF DUMMY VARIABLE FOR VAT FORECASTING OF THE TAX ADMINISTRATION IN THE SLOVAK REPUBLIC

Abstract

Prognozowanie jest ważną częścią działalności działu analizy wpływów podatkowych w Tax Directorate Ministerstwa Finansów Republiki Słowackiej w Banskiej Bystrzycy. Celem artykułu jest pokazanie jak modele SARMA, ARCH, GARCH mogłyby zostać użyte do krótkoterminowego prognozowania w miesięcznych cyklach. Prognozowanie jest konieczne dla skarbu pastwa. Według naszych przewidywa, skarb pastwa planuje krótkoterminową zagregowaną płynność na poziomie kraju. Podatkowa reforma w 2004, niepokojąco skomplikowała prognozowanie. Jest to powodem, dlaczego zastosowaliśmy korektę z użyciem dodatkowej zmiennej, która jest zaprezentowana w niniejszym artykule. To samo skorygowanie zostało wprowadzone w życie do modelu podatku VAT i do wszystkich modeli użytych w Tax Directorate. Modele przedstawione w artykule są próbką modeli rozwiniętych w dziale analizy wpływów podatkowych.

Key words

Dummy variable, SARMA models, ARCH models, GARCH models, VAT

1. Introduction

The State Treasury was established by State Treasury Act in 2002. The Act ordered the Tax Directorate of the Slovak Republic to send the tax forecasts to the State Treasury. The act does not specify the form how forecasts should be sent. We send forecasts on a monthly basis for the next three months. The closest month is forecast per day. The aim of the article is to show how we make monthly forecasts.

The fact that tax forecasts are made on a monthly basis influenced the explanatory variables. Classical macroeconomic variables cannot be used, because they are not observed on a monthly basis.

Moreover, another very important fact that influenced the forecasts is the Tax Reform from 2004 and the acceptance of the Slovak Republic for a member of the European Union.

2. VAT – theoretical knowledge

The movement from direct taxes to indirect taxes is obvious in the Slovak Republic. Indirect tax is also VAT.

VAT Act has gone through the many changes. We can notice the changes just from the year 1996, because of available data set. The most important change happened in 2004.

We suppose that the most important impacts on the tax revenue are: tax rate, taxation term, acceptance of the Slovak Republic for a member of European Union, tax reform (2004).

VAT Act recognises two kinds of taxation terms: Monthly taxation term and quarterly taxation term. We used dummy variable to describe the fact. Dummy variable gets figure 1 for months, where we apply quarterly terms (January, April, July, and October) and figure 0 otherwise. We also tested different forms of models, where we used average rate of the tax rate instead of 1 figure. Average rate of the tax rate were set as a weighed arithmetic mean. Weights were set from the aggregate VAT declaration. (There used to be two VAT rates valid at the same time in the Slovak Republic)

Tax rate is another important factor, which influences VAT revenue. Tax reform which occurred in 2004 brought just one 19 percent VAT rate. We used weighted arithmetic mean for the period before 2004. Table 1 describes VAT rates valid in the Slovak Republic.

Table 1. Basic and reduced VAT rate.

Year	Months	Reduced rate %	Basic rate %
1996	1	6	25
1996	2	6	23
1999	8	10	23
2003	2	14	20
2004	2	19	19

Source: Act No 289/1995 and No 222/2004 about VAT

According to the Table 1 and aggregated VAT declaration we calculated average VAT rate.

Acceptance of the Slovak Republic for the member of European Union influenced VAT revenues dramatically. Acceptance of the Slovak Republic led to the transfer of payments from the Customs administration to the Tax administration. The impact was noticed in July 2004. To model the impact, we developed a transformation which improved the forecasts. The transformation will be described in part 3.

3. Transformation developed at the Department of Tax Analysis

Acceptance of the Slovak Republic for a member of European Union influenced tax revenues very much. We modelled the fact using three versions of the dummy variable:

1) The dummy variable gets figure 0 from January 1996 – June 2004 and figure 1 from July 2004 – till now (01).
2) The dummy variable gets figure 1 from July 2004 – June 2005 and figure 0 otherwise because we found out that beginning July 2005 variable y_{t-12} carries also the information about the above mentioned impact (010).
3) The dummy variable gets figure 1 from July 2004 – June 2005, figure 0 from January 1996 – June 2004 and a figure which is neither 0 nor 1 for the beginning July 2005. The exact figure is a result of transformation developed at the Tax Analysis department.

Suppose VAT model in the simple form:

$$y_t = \beta_0 + \beta_1 y_{t-12} + \beta_2 y_{t-24} + \gamma x_t + \varepsilon_t + \theta_1 \varepsilon_{t-12} \tag{1}$$

where

$E(\varepsilon_t) = 0$

$E(\varepsilon_t^2) = \sigma^2$

$E(\varepsilon_t \varepsilon_\tau) = 0$

Where ε_{t-12} is moving average variable, x_t is a dummy variable and γ is a coefficient in million Slovak Crowns which describes the impact of the changes in 2004. x_t gets the figure 1 from July 2004 – June 2005.

The problem is, what happens in July 2005, which is the 13th month after the impact of the acceptance of the Slovak Republic for a European Union member.

VAT revenue y_{13} is

$$y_{13_2} = \beta_0 + \beta_1(\beta_0 + \beta_1 y_{t-24} + \beta_2 y_{t-36} + \theta_1 \varepsilon_{t-24} + \gamma + \varepsilon_{t-12}) + \beta_2 y_{t-24} + \gamma x_t + \varepsilon_t + \theta_1 \varepsilon_{t-12} \tag{2}$$

Equation (2.1) was rewritten to (2.2) where y_{t-12} was replaced by SARMA model without x_t variable. In case of using x_t variable, it should contain figure 1, in order to describe impact of the acceptance Slovak Republic for a European Union member. As the SARMA model, which replaced y_{t-12} does not contain x_t variable, it describes y_{t-12} as Slovak

Republic was not member of European Union. As y_{13_1} should describe the VAT revenue, we must add γ at the end.
Equation (2.1) can be also rewritten to equation (2.3):

$$y_{13_2} = \beta_0 + \beta_1(\beta_0 + \beta_1 y_{t-24} + \beta_2 y_{t-36} + \gamma + \varepsilon_{t-12} + \theta_1\varepsilon_{t-24}) + \beta_2 y_{t-24} + \gamma x_t + \varepsilon_t + \theta_1\varepsilon_{t-12} \quad (3)$$

In equation (2.3), y_{t-12} was replaced by SARMA model, which contains correction in terms of acceptance Slovak Republic for a European Union member. We used coefficient γ to achieve this. We did not use x_t there. We used x_t at the position where we expect neither 0 figure nor 1 figure, which is at the end of the equation (2.3). It is obvious that $y_{13_1} = y_{13_2}$, in consequence $x_{13} = 1 - \beta_1$. After the acceptance of the Slovak Republic for a European Union member, the x_t variable gets value $1 - \beta_1$ for months from 13th to 24th.

A change happens in 25th month, where y_{t-24} also contains information about the acceptance of the Slovak Republic for European Union member. Equations (2.4) and (2.5) are valid for the 25th months

$$\begin{aligned} y_{251} &= \beta_0 + \beta_1(\beta_0 + \beta_1(\beta_0 + \beta_1 y_{t-36} + \beta_2 y_{t-48} + \varepsilon_{t-24} + \theta_1\varepsilon_{t-36}) + \beta_2 y_{t-36} + \varepsilon_{t-12} + \theta_1\varepsilon_{t-24}) + \\ &+ \beta_2(\beta_0 + \beta_1 y_{t-36} + \beta_2 y_{t-48} + \varepsilon_{t-24} + \theta_1\varepsilon_{t-36}) + \gamma + \varepsilon_t + \theta_1\varepsilon_{t-12} \end{aligned} \quad (4)$$

$$\begin{aligned} y_{25_2} &= \beta_0 + \beta_1(\beta_0 + \beta_1 y_{t-24} + \beta_2 y_{t-36} + \gamma(1-\beta_1) + \varepsilon_{t-12} + \theta_1\varepsilon_{t-24}) + \beta_2(\beta_0 + \beta_1 y_{t-36} + \\ &\beta_2 y_{t-48} + \gamma + \varepsilon_{t-24} + \theta_1\varepsilon_{t-36}) + \gamma x_t \\ y_{25_2} &= \beta_0 + \beta_1(\beta_0 + \beta_1(\beta_0 + \beta_1 y_{t-36} + \beta_2 y_{t-48} + \gamma + \varepsilon_{t-24} + \theta_1\varepsilon_{t-36}) + \beta_2 y_{t-36} + \gamma(1-\beta_1) + \varepsilon_{t-12} + \theta_1\varepsilon_{t-24}) \\ &+ \beta_2(\beta_0 + \beta_1 y_{t-36} + \beta_2 y_{t-48} + \gamma + \varepsilon_{t-24} + \theta_1\varepsilon_{t-36}) + \gamma x_t \end{aligned} \quad (5)$$

It is obvious that $y_{25_1} = y_{25_2}$

$$\begin{aligned} \gamma &= \beta_1^2\gamma + \beta_1\gamma - \beta_1^2\gamma + \beta_2\gamma + \gamma x_t \\ x_t &= 1 - \beta_1 - \beta_2 \end{aligned} \quad (6)$$

We can see that x_t contains values which are the result of the computation process. We solved the problem in the iterative process. We used convergence criteria $|\beta_i - \beta_j| < 0.0001$. If the convergence is not achieved within 100 iterations, the iteration process stops.

As the transformation was developed at the Tax Analysis Department, we should test it for different models and for different forecast timeline.
We will test model (7) first:

$$y_t = \beta_0 + \beta_1 y_{t-12} + \gamma x_t + \varepsilon_t + \theta_1\varepsilon_{t-12} \quad (7)$$

We will explore model (2.7) for x_t variable for $t \in <1,12>, t \in <13,24>, t \in <25,36>$.
For $t \in <1,12>$ x_t gets value 1
For $t \in <13,24>$ x_t gets $1 - \beta_1$, what we can prove by:

$$\begin{aligned} y_t &= \beta_0 + \beta_1 y_{t-12} + \gamma x_t + \varepsilon_t + \theta_1\varepsilon_{t-12} \\ y_{t1} &= \beta_0 + \beta_1(\beta_0 + \beta_1 y_{t-24} + \varepsilon_{t-12} + \theta_1\varepsilon_{t-24}) + \gamma + \varepsilon_t + \theta_1\varepsilon_{t-12} \\ y_{t2} &= \beta_0 + \beta_1(\beta_0 + \beta_1 y_{t-24} + \gamma + \varepsilon_{t-12} + \theta_1\varepsilon_{t-24}) + \gamma x_t + \varepsilon_t + \theta_1\varepsilon_{t-12} \end{aligned}$$

$y_{t1} = y_{t2}$
$\gamma = \beta_1\gamma + \gamma x_t$
$x_t = 1 - \beta_1$
For $t \in <25,36>$ x_t gets $1-\beta_1$, what we can prove by:
$y_t = \beta_0 + \beta_1 y_{t-12} + \gamma x_t + \varepsilon_t + \theta_1\varepsilon_{t-12}$
$y_{t1} = \beta_0 + \beta_1(\beta_0 + \beta_1(\beta_0 + \beta_1 y_{t-36} + \varepsilon_{t-24} + \theta_1\varepsilon_{t-36}) + \varepsilon_{t-12} + \theta_1\varepsilon_{t-24}) + \gamma + \varepsilon_t + \theta_1\varepsilon_{t-12}$
$y_{t2} = \beta_0 + \beta_1(\beta_0 + \beta_1 y_{t-24} + \gamma(1-\beta_1) + \varepsilon_{t-12} + \theta_1\varepsilon_{t-24}) + \gamma x_t + \varepsilon_t + \theta_1\varepsilon_{t-12}$
$y_{t2} = \beta_0 + \beta_1(\beta_0 + \beta_1(\beta_0 + \beta_1 y_{t-36} + \gamma + \varepsilon_{t-24} + \theta_1\varepsilon_{t-36}) + \gamma(1-\beta_1) + \varepsilon_{t-12} + \theta_1\varepsilon_{t-24}) + \gamma x_t + \varepsilon_t + \theta_1\varepsilon_{t-12}$
$y_{t1} = y_{t2}$
$\gamma = \beta_1^2\gamma + \beta_1\gamma - \beta_1^2\gamma + \gamma x_t$
$\gamma = \beta_1\gamma + \gamma x_t$
$x_t = 1 - \beta_1$
For $t \in <37,48>$ x_t gets $1-\beta_1$, what we can prove by :
$y_{t1} = \beta_0 + \beta_1(\beta_0 + \beta_1(\beta_0 + \beta_1(\beta_0 + \beta_1 y_{t-48} + \varepsilon_{t-36} + \theta_1\varepsilon_{t-48}) + \varepsilon_{t-24} + \theta_1\varepsilon_{t-36}) + \varepsilon_{t-12} + \theta_1\varepsilon_{t-24}) +$
$\gamma + \varepsilon_t + \theta_1\varepsilon_{t-12}$
$y_{t2} = \beta_0 + \beta_1(\beta_0 + \beta_1(\beta_0 + \beta_1(\beta_0 + \beta_1 y_{t-48} + \varepsilon_{t-36} + \theta_1\varepsilon_{t-48} + \gamma) + \gamma(1-\beta_1) + \varepsilon_{t-24} + \theta_1\varepsilon_{t-36})$
$+ \gamma(1-\beta_1) + \varepsilon_{t-12} + \theta_1\varepsilon_{t-24}) + \gamma x_t + \varepsilon_t + \theta_1\varepsilon_{t-12}$
$y_{t1} = y_{t2}$
$\gamma = \beta_1^3\gamma + \beta_1^2\gamma - \beta_1^3\gamma + \beta_1\gamma - \beta_1^2\gamma + \gamma x_t$
$\gamma = \beta_1\gamma + \gamma x_t$
$x_t = 1 - \beta_1$

It is obvious that there are no changes in the variable x_t, after variable y_{t-12} contains information about the influence of acceptance of SR for the member of European Union,.

Another model, which will be tested is model (1). We have already explored $t \in <13,24>, t \in <25,36>$. We expect that x_t variable will get the same value in $t \in <37,48>$ as in $t \in <25,36>$. We can prove it by :

$$y_{t1} = \beta_0 + \beta_1(\beta_0 + \beta_1(\beta_0 + \beta_1(\beta_0 + \beta_1 y_{t-48} + \beta_2 y_{t-60} + \varepsilon_{t-36} + \theta_1\varepsilon_{t-48}) + \beta_2 y_{t-48} + \theta_1\varepsilon_{t-36} + \varepsilon_{t-24}) + \beta_2(\beta_0 + \beta_1 y_{t-48} + \beta_2 y_{t-60} + \varepsilon_{t-36} + \theta_1\varepsilon_{t-48}) + \theta_1\varepsilon_{t-24} + \varepsilon_{t-12}) + \beta_2(\beta_0 + \beta_1(\beta_0 + \beta_1 y_{t-48} + \beta_2 y_{t-60} + \varepsilon_{t-36} + \theta_1\varepsilon_{t-48}) + \beta_2 y_{t-48} + \theta_1\varepsilon_{t-36} + \varepsilon_{t-24}) + \theta_1\varepsilon_{t-12} + \gamma + \varepsilon_t \quad (8)$$

$$y_{t2} = \beta_0 + \beta_1(\beta_0 + \beta_1(\beta_0 + \beta_1(\beta_0 + \beta_1 y_{t-48} + \beta_2 y_{t-60} + \gamma + \varepsilon_{t-36} + \theta_1\varepsilon_{t-48}) + \beta_2 y_{t-48} + \gamma(1-\beta_1) + \varepsilon_{t-24} + \theta_1\varepsilon_{t-36}) + \beta_2(\beta_0 + \beta_1 y_{t-48} + \beta_2 y_{t-60} + \gamma + \varepsilon_{t-36} + \theta_1\varepsilon_{t-48}) + \gamma(1-\beta_1-\beta_2) + \varepsilon_{t-12} + \theta_1\varepsilon_{t-24}) + \beta_2(\beta_0 + \beta_1(\beta_0 + \beta_1 y_{t-48} + \beta_2 y_{t-60} + \gamma + \varepsilon_{t-36} + \theta_1\varepsilon_{t-48}) + \beta_2 y_{t-48} + \gamma(1-\beta_1) + \varepsilon_{t-24} + \theta_1\varepsilon_{t-36}) + \gamma x_t \quad (9)$$

$y_{t1} = y_{t2}$
$\gamma = \beta_1^3\gamma + \beta_1^2\gamma - \beta_1^3\gamma + \beta_1\beta_2\gamma + \beta_1\gamma - \beta_1^2\gamma - \beta_1\beta_2\gamma + \beta_2\beta_1\gamma + \beta_2\gamma - \beta_1\beta_2\gamma + \gamma x_t$
$\gamma = \beta_1\gamma + \beta_2\gamma + \gamma x_t$
$x_t = 1 - \beta_1 - \beta_2$

It is obvious that there are no changes in the variable x_t, after y_{t-12} and y_{t-24} contain information about the acceptance the Slovak Republic for the member of European Union,.

4. How we tested

We tested 15 540 models of VAT revenue. We used many variables. Unfortunately, we could not use the y_{t-1} variable, although variable y_{t-1} has its economic interpretation. It is because of technical reasons. The deadline for sending forecast is the 20th in a month, which means, that we actually do not know y_{t-1} observation when we make forecasts. We considered just the variables $y_{t-3}, y_{t-12}, y_{t-24}$.

We tried to find the best model which would replace present model. The most important criteria was the quality of the ex post analysis. The quality of the ex post analysis was measured by RMS (Root mean square).

$$RMS = \sqrt{\left(\frac{1}{n}\sum_{t=1}^{n}\left(y_t^s - y_t\right)^2\right)} \quad (10)$$

Root mean square has a very important property, it penalises larger deviations from smaller.

We carried out ex post analysis exactly in the same way, how it would work in real situation. State Treasury wants us to send them 3 months forecast and the most important month is the first forecasted month. We send forecasts in those months, which we do not know the real figures. This is the reason, why the data basis is set 2 months back compared with the forecasted months. We ran ex post analysis from January 1996 to December 2006 on the data basis set 2 months back compared with the forecasted months.

$$y_t = \beta_0 + \beta_1 y_{t-3} + \varepsilon_t \quad (11)$$
$$y_t = \beta_0 + \beta_1 y_{t-12} + \varepsilon_t \quad (12)$$
$$y_t = \beta_0 + \beta_1 y_{t-3} + \beta_2 y_{t-12} + \varepsilon_t \quad (13)$$
$$y_t = \beta_0 + \beta_1 y_{t-3} + \beta_2 y_{t-12} + \beta_3 y_{t-24} + \varepsilon_t \quad (14)$$
$$y_t = \beta_0 + \beta_1 y_{t-12} + \beta_2 y_{t-24} + \varepsilon_t \quad (15)$$
$$y_t = \beta_0 + \beta_1 y_{t-3} + \varepsilon_t + \theta_1 \varepsilon_{t-12} \quad (16)$$
$$y_t = \beta_0 + \beta_1 y_{t-12} + \varepsilon_t + \theta_1 \varepsilon_{t-12} \quad (17)$$
$$y_t = \beta_0 + \beta_1 y_{t-3} + \beta_2 y_{t-12} + \varepsilon_t + \theta_1 \varepsilon_{t-12} \quad (18)$$
$$y_t = \beta_0 + \beta_1 y_{t-3} + \beta_2 y_{t-12} + \beta_3 y_{t-24} + \varepsilon_t + \theta_1 \varepsilon_{t-12} \quad (19)$$
$$y_t = \beta_0 + \beta_1 y_{t-12} + \beta_2 y_{t-24} + \varepsilon_t + \theta_1 \varepsilon_{t-12} \quad (20)$$
$$y_t = \beta_0 + \beta_1 y_{t-3} + \varepsilon_t + \theta_1 \varepsilon_{t-12} + \theta_2 \varepsilon_{t-24} \quad (21)$$
$$y_t = \beta_0 + \beta_1 y_{t-12} + \varepsilon_t + \theta_1 \varepsilon_{t-12} + \theta_2 \varepsilon_{t-24} \quad (22)$$
$$y_t = \beta_0 + \beta_1 y_{t-3} + \beta_2 y_{t-12} + \varepsilon_t + \theta_1 \varepsilon_{t-12} + \theta_2 \varepsilon_{t-24} \quad (23)$$
$$y_t = \beta_0 + \beta_1 y_{t-3} + \beta_2 y_{t-12} + \beta_3 y_{t-24} + \varepsilon_t + \theta_1 \varepsilon_{t-12} + \theta_2 \varepsilon_{t-24} \quad (24)$$
$$y_t = \beta_0 + \beta_1 y_{t-12} + \beta_2 y_{t-24} + \varepsilon_t + \theta_1 \varepsilon_{t-12} + \theta_2 \varepsilon_{t-24} \quad (25)$$

We tested models (11) – (25). Models (11) – (25) were also tested in versions without intercept.

We used following models: SARMA, ARCH(1), ARCH(2), GARCH(1,1), GARCH(2,1), GARCH(2,2).

We applied transformation developed by us in two different ways :

1) We applied ARCH and GARCH models from the beginning of iterations (transf).
2) We applied SARMA models until convergence was achieved and ARCH or GARCH models were applied just in the final iteration (transf2).

The dummy variable, which describes acceptance of the Slovak Republic for a member of European Union was used in all four described versions (01, 010, transf, transf2).

We added from 1 to 3 variables. The dummy variable which describes the acceptance of the Slovak Republic for a member of the European Union has never been omitted (x_{t1}).

We used both raw data and logarithmised data.

The variable x_{t2} described the influence of the quarter year tax payers. The influence of the quarter year tax payers was described by the dummy variable which got figure 1 in January, April, July, October and figure 0 otherwise (0,1,0). We used also different version of the dummy variable, where we used weighed average rate instead of figure 1 (Rates).

The variable x_{t3} described the influence of the tax rate. The influence of the rate was described by the weighed average mean of the monthly tax payers (MTP) or with the weighed average mean of the monthly tax payers and the quarterly tax payers (MQTP).

Table 2. The smallest RMS of the VAT models with the variables x_{t1}, x_{t2}, x_{t3}

Nr.	Nr.of model	Model	The Dummy Variable	Log Yes/No	RMS	Variables	Intercept
1	(3.11)	GARCH(1,1)	transf2	Yes	803	MTP	Yes
2	(3.11)	GARCH(1,1)	transf	Yes	843	MTP	Yes
3	(3.11)	ARCH(1)	010	Yes	845	MTP	Yes
4	(3.11)	SARMA	transf2	Yes	850	MTP	Yes
5	(3.11)	GARCH(1,1)	010	No	858	-------	Yes
6	(3.11)	GARCH(1,1)	010	Yes	867	MTP	Yes
7	(3.11)	SARMA	transf	Yes	869	MTP	Yes
8	(3.11)	ARCH(1)	transf2	Yes	874	MTP	Yes
9	(3.8)	ARCH(1)	transf2	Yes	884	(0,1,0),MTP	No
10	(3.8)	ARCH(1)	transf2	Yes	887	Rates, MTP	No

Source: own calculations using data from the Tax Directorate of the Slovak Republic.

Table 2 shows 10 best models with the smallest RMS. We ran ex post analysis from January 2005 to December 2006. The best model in the Table Nr. 2 is model (26):

$$y_t = \beta_0 + \beta_1 y_{t-12} + \gamma_1 x_{t1} + \gamma_2 x_{t2} + \varepsilon_t + \theta_1 \varepsilon_{t-12} \tag{26}$$
$$\sigma_t^2 = \omega + \alpha_1 \varepsilon_{t-1}^2 + \delta_1 \sigma_{t-1}^2$$

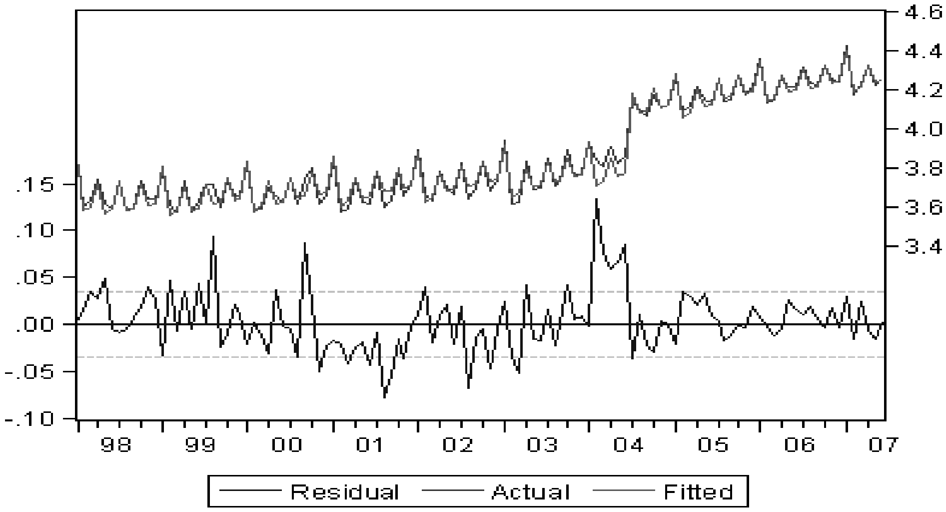

Fig. 1 Graph of the actual values, fitted values and residuals of the model (26)
Source: Own calculations using data from the Tax Directorate of the Slovak Republic.
Table 3. Model (26) – Eviews 6 output.

	Coefficient	Std. Error	z-Statistic	Prob.
C	-0.276	0.140	-1.973	0.0484
Y(-12)	0.945	0.096	9.831	0
Y(-24)	0.029	0.100	0.292	0.77
X2	0.299	0.027	10.883	0
X1	0.022	0.008	2.614	0.0089
MA(12)	-0.943	0.037	-24.828	0
Variance Equation				
C	0.0008	0.0005	1.558	0.119
RESID(-1)^2	0.189	0.145	1.297	0.194
GARCH(-1)	0.369	0.335	1.101	0.270
R-squared	0.982	Mean dependent var		3.861
Adjusted R-squared	0.980	S.D. dependent var		0.250
S.E. of regression	0.034	Akaike info criterion		-3.787
Sum squared resid	0.125	Schwarz criterion		-3.571
Log likelihood	224.882	Hannan-Quinn criter.		-3.699
F-statistic	729.757	Durbin-Watson stat		1.561
Prob(F-statistic)	0.000			

Source : Own calculations using data from the Tax Directorate of the Slovak Republic.

Table 4. Autocorrelation and partial correlation function of the residuals (26).

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.007	-0.007	0.0056	
		2	0.009	0.009	0.0158	0.900
		3	0.007	0.008	0.0225	0.989
		4	0.118	0.118	1.7042	0.636
		5	-0.060	-0.059	2.1428	0.710
		6	-0.018	-0.022	2.1847	0.823
		7	-0.044	-0.046	2.4232	0.877
		8	-0.033	-0.047	2.5580	0.923
		9	-0.019	-0.004	2.6029	0.957
		10	-0.067	-0.066	3.1778	0.957
		11	0.078	0.089	3.9590	0.949
		12	0.063	0.071	4.4705	0.954
		13	0.032	0.030	4.6039	0.970
		14	-0.054	-0.047	4.9962	0.975
		15	0.116	0.085	6.8047	0.942
		16	-0.095	-0.111	8.0134	0.923
		17	-0.050	-0.059	8.3521	0.938
		18	0.076	0.101	9.1525	0.935
		19	-0.018	-0.036	9.1990	0.955
		20	-0.062	-0.022	9.7351	0.959
		21	-0.052	-0.039	10.118	0.966
		22	-0.024	-0.046	10.197	0.976
		23	-0.035	-0.032	10.375	0.983
		24	0.060	0.050	10.913	0.984

Source: Own calculations using data from the Tax Directorate of the Slovak Republic.

Table 2 shows that the influence of the quarter tax payers is not significant. Model with variable which describes quarter tax payer appears on the position No 9.

Table 4 shows that there is no serial correlation in the model (26)

Q statistics values are calculated according to Ljung – Box test of serial correlation, which is modified version of Box – Pierce test. Ljung – Box test is defined as :

$$Q = n(n+2)\sum_{i=1}^{p} \frac{\hat{\rho}_i^2}{n-i} \qquad (27)$$

where

$$\hat{\rho}_i = \frac{\operatorname{cov}(e_t, e_{t-i})}{\sqrt{\operatorname{var}(e_t)}.\sqrt{\operatorname{var}(e_{t-i})}} \quad \text{for i=1,2,3,.......,p} \qquad (28)$$

Q statistics has χ^2 distribution with p degrees of freedom.

Stationarity is also very important. We tested stationarity with the roots of the equation (29)

$$1-(\delta_1+\alpha_1)z = 0 \qquad (29)$$

Process is stable, when roots of the equation (29) lie outside the unit circle. $\alpha_1 = 0{,}189$ and $\delta_1 = 0{,}369$ for model (26). Process is stable. Stationarity causes that influence of the random errors dies out in t+j. The effect is stronger with growing j.

Table 5 shows specific values of the VAT from January 2005 to December 2006.

Table 5. Ex post forecast 2005 – 2006. In million Sk

Month	2005		2006	
	Reality	(3.17)	Reality	(3.17)
1	18 082	18 253	22 599	22 129
2	12 336	11 610	13 577	13 974
3	12 858	11 564	13 881	15 017
4	16 468	15 893	18 323	19 481
5	13 904	13 556	16 261	15 998
6	13 476	13 694	16 279	15 906
7	18 031	17 234	20 516	20 262
8	13 709	14 407	16 880	15 443
9	14 048	12 711	16 473	15 955
10	18 619	17 958	20 814	19 888
11	14 945	14 937	17 656	16 416
12	16 169	15 004	17 308	17 411

Source: Own calculations using data from the Tax Directorate of the Slovak Republic.

5. Ex post and ex ante forecast in 2007.

Table 6 shows specific values of the ex post forecast from January 2007 to June 2007. Column *ex post* 1 means that forecast was created on the data basis from January 1996 to December 2006. *Ex post* 2 means that forecast was created on the data basis set 2 months back compared with the forecasted months.

Table 6. VAT ex post forecast from January 2007 to June 2007 in million Sk

2007					
VAT	ex post 1	ex post 2	real	For./Real. 1	For./Real. 2
1	22 801	22 305	26 568	0,86	0,84
2	15 527	15 497	14 932	1,04	1,04
3	15 688	15 644	17 108	0,92	0,91
4	20 009	20 542	20 604	0,97	1,00
5	17 563	17 927	16 682	1,05	1,07
6	17 596	17 484	17 831	0,99	0,98

Source: Own calculations using data from the Tax Directorate of the Slovak Republic.

Table 7 shows specific values of the ex ante forecast from July 2007 to December 2007. Ex ante 1 means that forecast was created on the data basis from January 1996 to December 2006. Ex ante 2 means that forecast was created on the data basis from January 1996 to June 2007.

Table 7. Ex ante forecast from July to December 2007 in million Sk

2007		
Month	ex ante 1	ex ante 2
7	21 132	22 530
8	17 646	18 181
9	17 749	18 345
10	21 960	23 498
11	18 458	19 154
12	18 895	19 702

Source : Own calculations using data from the Tax Directorate of the Slovak Republic.

6. Conclusion

We described how Tax Directorate of the Slovak Republic could forecast VAT. We also described transformation developed at the Tax Analysis Department.

We suppose that the transformation could be applied in econometric modelling. As ex post analysis showed, transformation improved forecasts. Its advantage is that forecasts are less unbiased than forecasts prepared with using just bivalent form of the dummy variable, but the assumption of deterministic change in July 2004 is simplified.

We used SARMA models, ARCH models and GARCH models. The reason for using ARCH and GARCH models was simple. It was an experiment. We tried to find out, whether ARCH and GARCH models improve forecasts or not. The experiment was successful and ARCH and GARCH models improved forecasts. As we know, ARCH LM test tests the existence of the conditional heteroscedasticity. We didn‘t use it and we just tested the relevance of the ex post analysis measured with RMS. RMS and economic interpretation of the variables were most important factors which determined the quality of models. Stationarity and serial correlation were tested afterwards.

ARCH and GARCH models were designed for financial data short-term modelling (on the daily basis or even shorter) and as we could see, they can be used on the monthly basis, too.

We used approach which is somewhere between econometrics and statistics. We borrowed data mining from statistics and we used just variables with economic interpretation (econometrics).

The main disadvantage of the presented models, described transformation and procedure is its consumption of computer performance. Two (or better four) core processor is necessary. Another disadvantage which hasn’t been fully explored is the convergence of the transformation. We discovered that convergence was not achieved in many cases.

The main advantage is the precision of the forecasts. Reviewing of the article is a process which lasts some time. We already know the results of the ex ante forecasts. The results are good. December 2007 is a 100 % forecast (99,6 %) which is very good result and it is made on the data basis January 1996 – June 2007 which is 6 observations beyond last observation.

Table 8. Forecasts in million Sk

VAT	Forecast			for./real.	for./real.
2007	ex ante 1	ex ante 2	Reality	1	2
7	21 132	22 530	21 893	0,97	1,03
8	17 646	18 181	17 248	1,02	1,05
9	17 749	18 345	17 381	1,02	1,06
10	21 960	23 498	22 961	0,96	1,02
11	18 458	19 154	20 663	0,89	0,93
12	18 895	19 702	19 787	0,95	1,00
$\sum$	115 840	121 410	119 933	0,97	1,01

Source : Own calculations using data from the Tax Directorate of the Slovak Republic.

Literature

[1] ARLT, J., ARLTOVÁ, M. : Finanční časové řady. Praha : Grada, 2003

[2] ARLT, J., ARLTOVÁ, M. : Ekonomické časové řady. Praha : Grada, 2006

[3] DAVIDSON, R., MACKINNON, J. G. : Estimation and Inference in Econometrics. New York, Oxford University Press, 1993

[4] GRANGER, C. W. J. : Forecasting in business and economics. San Diego : Academic Press, 1989

[5] HAMILTON, J.D. : Time series analysis. Princeton : Princeton 1994

[6] HATRÁK, M. : Ekonometrické metódy I. Bratislava : Ekonóm, 1995

[7] HATRÁK, M. : Ekonometrické metódy II. Bratislava : Ekonóm, 1995

[8] HAYASHI, F. : Econometrics. Princeton : Princeton University Press 2000

[9] VOGELVANG, B. : Econometrics Theory and Applications with Eviews : Financial Times Prentice Hall, 2005

[10] Act Nr. 289/1995 Z.z. o dani z pridanej hodnoty

[11] Act Nr. 222/2004 Z.z. o dani z pridanej hodnoty

[12] Act Nr. 291/2002 Z.z. o Štátnej pokladnici

[13] Eviews 5.1 Command and programming reference

[14] LUKÁČIK, M., LUKÁČIKOVÁ, A., SZOMOLÁNYI, K.: Ekonometrické prognózovanie importu SR. Bratislava : Ekonomické rozhľady 2/2007

[15] KRAJČÍR, Z., KRÁLOVÁ, J., LIVERMORE, S. : Prognózovanie dane z pridanej hodnoty v SR, www.finance.gov.sk

[16] http://jaspi.justice.gov.sk

[17] www.drsr.sk

[18] www.mpavlik.net