Econometrical Modelling of Profit Tax Revenue

R. Rudzkis[1], E. Mačiulaitytė[2]

[1]Institute of Mathematics and informatics, Akademijos str. 4, LT-08663 Vilnius
rudzkis@ktl.mii.lt
[2]Vilnius Gediminas technical university, Saulėtekio ave. 11, LT-10223 Vilnius
elemac@takas.lt

Abstract. The aim of this article is to present a forecast of budget revenue from the profit tax using econometric models. The set of applied models has to be reduced to very simple models due to short time series used. Therefore, the profit tax regression analysis is made in two stages. In the first stage, econometric modelling of profit tax revenue with the main profit indicators (called the profit tax base) is performed on the basis of information on profit tax regulation and its changes. In the second stage, algorithms of forecasting the profit tax base are formed when the main macroeconomic indicators of Lithuanian economy are used as regressors. Crossvalidation was applied to estimate the accuracy of these algorithms.

Keywords: the linear regression model, the error correction model, profit tax revenue profit.

1   Introduction

Accurate (exact) forecast of tax revenue is a very important task for state budget planning. Both underestimation and overestimation of planed revenue could cause problems when the revenue is used to finance government functions. The main and largest part of revenue is tax revenues which are collected from certain taxes. For the past four years revenue from the profit tax was growing faster as compared with other tax revenues. Therefore, its importance has grown up too: last year the share of profit tax revenue was 12 percent of all tax revenues. The past few years planning of profit tax revenue was not very exact: the sum of the profit tax collected was considerably smaller or considerably larger than the planned profit tax revenue. The difference between the actual and planned revenue was about 12–56 percent every year.

There are several related factors which complicate profit tax revenue modelling. It is doubtful if the indicator of profit tax revenue is stationary. The assumption of stationarity of indicators is usually made when applying econometric models to the indicators. This problem is caused by a frequent change of the Profit tax law. More thorough analysis of this problem is presented in the article by E. Mačiulaitytė [1]. In addition, taxation order

is defined in the acts of legislation, the basic tax elements determine the relations between indicators – they define a strict character of relationship.

The phenomenon of non-legal tax evasion and tax planning (that is a legal manipulation of income, cost, allowances, and other elements in order to lower the profit tax) is typical of the Lithuanian economy. About 20 percent of the estimated and officially announced shadow economy is the evidence of tax evasion. Some corrections of the profit tax law that could be called "patching of holes" as the means of legal tax lowering are annulled could be related to the phenomenon of tax planning.

In [1], the problems of modelling of the profit tax revenue indicator are suggested to solve by dividing the modelling into two stages: estimation of the dependence of profit tax revenue regressions on the main profit indicators of economy and formation of forecast algorithms of the profit tax base. The relationship between the tax revenue and the main macro indicators of economy is complicated therefore its approximation by models with only few parameters cannot ensure the required accuracy (of a model and forecast). However, the treatment of this relationship as a composition of two simpler functions and separate estimation of each of them facilitates the solution of the problem and enables us to apply fairly straightforward models. This is of great importance because transitional processes were typical of the Lithuanian economy in the past few years. Under these conditions fairly short time series could be used for statistical identification of models, because the statistical data of prior years cannot reflect the current economic structure. In addition, official state statistics on quarterly profit indicators has been published only since the beginning of 1998. Hence, the actual forecast of profit tax revenue is based on very small samples. The difficulties dealt with in this work are rather characteristic of the application of statistical methods in social and economic researches.

In the first part of the paper, the models of profit volume and profit tax indicator met in other sources are reviewed in short. In the second part of the paper, specification of profit tax revenue models and choice of alternative indicators for approximation of the true profit base are discussed. In the third and fourth parts of the paper, the models of Lithuanian profit indicators and regression functions of profit tax revenue are presented with accuracy estimations of the respective with certain forecasting algorithms.

2 The experience of profit and profit tax revenue modelling

There are many special works describing researches on profit modelling on the micro level, i.e. on the individual company level. The profit is usually defined as the difference between production output and production input which maximizes a single company. Various forms of production functions which are nonlinear with respect to their arguments are included into profit equations. The prices of output and prices of inputs, labour force, wages, capital stock, and material investment are analyzed as the main factors that influence the profit value. The company's profit maximization task is also determined by other factors such as demand of production, export, prices of import, etc.

The profit of the national economy is the sum of profits of individual companies; therefore it could be evaluated both by modelling macroeconomic relations and by ana-

lyzing those relations on the micro level. In the latter case, the activity model of the major part of companies are necessary; however, statistical data of the individual company activity indicators are not publicly supplied by official statistics. Governmental institutions (the State Tax Inspectorate, etc.) which have the disposition of the data base of individual company indicators could possibly use the results of micro level research to evaluate the profit of the whole economy; however it is hardly reasonable because of large labour expenditure. Besides, the instability of set of companies could cause additional difficulty, because every year a lot of new companies are established and many old ones terminate their activity, or a company is reorganized into several companies, or it establishes a daughter enterprise, or several companies are conjoined into one, etc.

Despite all problems, modelling on the micro level provides a lot of useful experience, which could be used for a specification of equations on the macro level: for choosing regressors, type of equations as well as for defining the lag for influence of factors. It is worth mentioning the work of S. Kumbhakar [2], where profit functions are derived under the assumption of technical inefficiency or allocative inefficiency, or both. In that article economic indicators are modeled in logarithmic form. J. Bradley et al. [3] emphasize possible nonlinear relations between profit and investment which was testified by the research of the authors of this article.

In this work, we are interested in modelling of profit only on macro level, i.e. on a state economy scale. Unfortunately, there are not many articles dedicated directly to this topic. The equation of total sum of profit of New Zealand companies [4] is constructed corresponding to the same relations on the micro level

$$B = P_D \cdot Y_D + P_E \cdot Y_E - W \cdot E + P_M \cdot M,$$

where B is profit, P_D are prices of domestic goods, P_E is the price of exports, P_M is the price of import of raw materials, Y_D is output of domestic goods, Y_E – exports, M denotes import of raw materials, W is the wage rate, and E denotes the private sector employment.

In macro econometric models of some countries (BOF5 [5, 6] of Finland, HERMIN HE4 [3] of Estonia) profit is defined as the difference between the gross domestic product (GDP) of the country and compensation of employees, consumption of the fixed capital, taxes on production and imports less subsidies. As the primary statistical analysis has shown, the profit indicator defined in this way has a week correlation with the profit tax revenue in Lithuania; therefore it is not suitable for the role of the profit tax base.

In the Lithuanian macro model LITMOD [7], profit is defined as the difference of income of production output and inputs (raw materials, wages of employees, investment), and the production output is described by nonlinear production functions. The set of equations of the macro model allows to define complex relations between profit and the main macroeconomic indicators. The LITMOD profit indicator could be used as one of the profit tax base approximations. Nevertheless, the profit cannot be easily predicted by means of equations of the model because prediction of output income and especially that of inputs is problematic; there is a lack of data of some input elements.

In the works aimed at profit tax revenue modelling the revenues $T(t)$ are usually

described by linear functions:

$$T(t) = \beta_0 + \beta_1 \cdot R(t) \cdot B(t) - A(t) + \varepsilon(t),$$

where $B(t)$ is profit, $A(t)$ are allowances, $R(t)$ is the tax rate, and $\varepsilon(t)$ is the error term of the model.

In some articles, nonlinear relations are pointed out. P. Van den Noord [8] shows the nonlinear dependence of the profit tax revenue on the wage rate or employment. S. Kennedy et al. [9] used logarithmic transformations and approximated the Australian profit tax revenue by means of error correction type of equation

$$\Delta T^*(t) = \alpha_0 + \alpha_1 \Delta R^*(t) + \alpha_2 \Delta B^*(t) + \alpha_3 \log T^*(t-1)$$
$$+ \alpha_4 B^*(t-1) + \alpha_5 R^*(t-1) + \alpha_6 D_1(t-1) + \alpha_7 D_2(t-1).$$

Here and in the sequel, X^* (index with an asterisk) denotes the logarithm of indicator X; Δ is the first-difference operator, D_1 and D_2 are dummy variables reflecting changes in Australian taxation (change-points).

Y. Hsing [10] assumes that there is nonlinear regression dependence of the tax revenue $T(t)$ on tax rate $R(t)$

$$T(t) = \beta_1 R(t) + \beta_2 r^2(t) + \varepsilon(t),$$

where the coefficient $\beta_2 < 0$ shows the influence of tax evasion, the higher the rate, the more incentive is not to show the earned income (profit).

Possible nonlinear relations between the profit tax revenue and the factors influencing profit as well as lack of data motivate dividing of profit tax revenue modelling into two stages mentioned in the introduction of the article in order to get a more accurate forecast using simple models. The Lithuanian profit tax revenue approximation using wide spread straightforward models of the main macroeconomic indicators applied by A. Budrytė and E. Mačiulaitytė [11] described the scattering of data fairly well, however the respective forecasts were not accurate enough. This testifies that in the Lithuanian case, more profound research is necessary as the statistical data are scarce and the economy was in transition (privatization and integration into the European Union (EU), etc. were in progress) during the reviewed period.

3 Peculiarities of model specification

As mentioned above, quarterly aggregated activity indicators of companies only from 1998 could be used for econometric analysis of Lithuanian budget profit tax revenue which means that the time series of statistical data available are very short. On the other hand, the revenue regression dependence on other macroeconomic indicators and various tax rates is rather complicated; therefore the adequacy of econometric models developed is mainly determined by a priori specification: a set of assumptions made on the base of the expert analysis of economic relations and the taxation system.

The main rules of profit tax regulation in Lithuania and their development are extensively described in the article by E. Mačiulaitytė [1], so only the most important aspects are mentioned in this work. During the considered period (since 1998), the profit tax rate has been changed two times: at the beginning of 2000 the rate was lowered from 29 % to 24 %, and from January of 2002 the rate of 15 % was established and the zero tax rate for profit, used for investment, was cancelled at the same time. From the first view, profit tax revenue $T(t)$ could be exactly approximated by the product $B(t)R(t)$

$$T(t) = B(t) \cdot R(t) + \varepsilon(t),$$

where $B(t)$ is the profit before tax (PBT) of companies which worked profitably during the period t, $R(t)$ is the tax rate. In reality, there is a more complicated relationship. Not so small part of profit tax revenue consists of profit tax advanced payments. For each company they are set by the profit of the past years (the last year and the year before last) or by the indicators of current year (dependent on the choice of company). In addition, if a company had losses in the past few years, it can deduct these losses from the profit (thus it reduces the profit tax base) when calculating the profit tax for the last year. It is allowed by the law to carry over the losses to the next year, five years forward. Therefore, in the case the profit tax of a company, eg. 2005, depends on the activity results during the period of 2000–2005 of the company.

Due to these and other reasons mentioned above, the primary statistical analysis does not show that profitably working companies PBT is a more informative regressor of revenue $T(t)$ in comparison with all nonfinance enterprises PBT (the difference between profit of profitable companies and losses of profitless companies) or that of all companies operating profit (loss). In the last part of this paper the revenue $T(t)$ regression dependence on the tax rate $R(t)$ and so-called tax base $B(t)$, which is approximated by each of three profit indicators mentioned above, is modelled. Since the ultimate aim of this work is the most accurate forecast of $T(t)$, the suitability of a certain profit indicator for the profit tax base role depends not only on the accuracy of fitted regression models

$$T(t) = f(B, R) + \varepsilon$$

but also on the quality of forecasting procedures constructed for the chosen tax base indicator $B(t)$.

Modelling of dependency of all the three mentioned indicators on the main macroeconomic indicators is presented in the next part of this paper. Specifying the pending econometric models, it is essential to take in consideration changes in business conditions (internal and external) during the period analyzed. Experts mark out two events with one accord: the Russian crisis in September of 1998 and Lithuanian integration into EU. The actual devaluation of the Russian rouble induced a negative shock to Lithuanian exports in 1999 which also affected the whole production. The recession of economy affected companies' activity indicators with a lag, i.e., the "bottom" of profitability was reached in 2000 consequently the appropriate "change-point" in the model should include the period from the second half of 1999 to the first half of 2000. It is much more complicated to comment the influence of EU expansion on the development of Lithuanian economy.

Formally EU extended in May of 2004 although it is doubtful if this date should be the appropriate change-point of the model since actually the integration of new members into EU was rather a long-term process than an instantaneous event. In Lithuania it began in 1995 and gained more action around 2002.

Such a feature of the Lithuanian economy as a large extent of shadow economy needs to be pointed out. The Department of Statistics and Lithuanian Free Market Institute estimated the extent of 20 %. Thus, tax evasion is rather spread in Lithuania, therefore the mentioned above Y. Hsing [10] assumption of nonlinear $T(t)$ dependence on $R(t)$ is noteworthy. The statistical data also corroborate this: the ratio of profit tax revenue and gross domestic product (GDP) in 2004–2005 was much higher than in 1997–1998 although the tax rate was almost two times lower and the economy was growing almost at the same rate during these two periods.

Even though the sales in the country and abroad increase a company's income their impact on profit indicators differ. When selling goods and services in the country the whole income is received at once, whereas a company that exports goods receives returned sums of the value added tax for purchased goods used for export production from the State Tax Inspectorate only after some time. On the other hand, to take a good position in foreign markets is much more difficult than in the local market, besides there are more opportunities to expand production when trading abroad. Therefore, at the beginning of expansion, in order to penetrate into larger markets, companies usually keep dumping prices and even export at a loss and putting aside plans of higher profitability into future, when solid relations with the partners becomes settled. Due to these mentioned reasons, an increase of export volume in fact positively affects the profit with a delay. We suggest here to approximate the average profit reaction to a single change in export by the type of functions presented in Fig. 1.
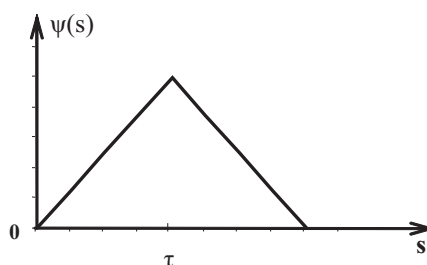

Fig. 1. The function of the profit reaction to the export change.

In Fig. 1. s denotes the time passed from the moment of export increase to the current moment, τ denotes the time point when the crosscorrelation function of profit and export changes reaches its maximum.

Because of a particular effect of export on profitability, the decomposition of the whole economy into two sub-sectors – the main exporting sectors (manufacturing, transport) are attributed to the first group and the rest ones – to the second one, – is noteworthy. The profit of different sub-sectors is described by different models and profit of the whole economy consists of their sum.

A complete specification of the model depends not only on the chosen type of model and statistical data available but also on the parameter estimation method applied. Along with the usual least squares estimation method, the median estimation method, which is less sensitive to deviant observations (i.e. more robust) was applied in this work.

4 Modelling of the profit indicators

Since this work is aimed at predicting the profit tax revenue, only those Lithuanian economy profit indicators are modelled which have a strong correlation with dependent variable $T(t)$. As mentioned in the previous section, there are three indicator expectants to approximate the profit tax base:

1) the profit of non-financial enterprises that worked profitable during the current period (PBT^+),

2) the difference between PBT^+ of all profitable non-financial enterprises and others losses (PBT),

3) the operating profit (loss) of all non-financial enterprises (OP).

In all the equations given below, the value of each of the three indicators at the moment t is denoted by $B(t)$ pointing out which of the indicators used in context is basic.

4.1 Regressors

In the econometrical modelling of profit rather complicated problem is selection of most informative regressors. In nowadays economic the profitability of enterprise is influenced by many internal and external factors the impact of which is rather often unstable in the country with transitional economy such as Lithuania. In addition, the lack of statistical data (short time series) prevents from including a great number of factors into the model equations, therefore only several main factors are considered. When selecting the main factors, it is essential to take into account both the statistical criteria and the laws of economic theory and conclusions of economic experts.

As the authors had predicted in advance, the preliminary statistical analysis has shown strong dependence of profit on GDP (it is denoted by Y in the models), material investment (I) and wage fund (F). In some equations (models), the significant regressors were export volume (X), productivity (λ), and the bank credit average interest rate (r). Even though the flow of foreign direct investment which promotes sales and export growth [12] undoubtedly has an indirect influence on the profit indicators, its direct impact on the profitability of economy statistically was not significant. Similar conclusions have been drawn when analysing the influence of the unemployment rate, the price indexes and import volume on the profit. The latter indicators have not been included into the profit models presented here mainly because of the small data sample. In the future having longer data time series, the set of regressors can be enlarged. Since monthly profit indicators are not (registered and) published, the quarterly indicators are analyzed, i.e. the unit of time is one quarter.

When estimating the export impact on profit, the moving average of export changes has been used as the regressor in this work

$$\widetilde{X}(t) = \sum_{s=0}^{2\tau} \left(1 - \frac{|s - \tau|}{\tau}\right) \Delta X(t - s),\tag{1}$$

where τ is the time point, where the cross-correlation function $r(\tau)=\mathrm{cor}(\Delta B(t), \Delta X(t-\tau))$ reaches its maximum. The statistical analysis has shown that the strongest correlation is after 3 quarters, i.e. $\tau = 3$.

The equations for profit calculation on the micro-level show that it is expedient to use moving averages for modelling the impact of material investment on profit. An enterprise operating profit is approximately calculated as a difference between the enterprise value added and compensation of employees (wages and social contributions) as well as depreciation of the accumulated capital. The latter can be expressed as a weighted sum:

$$\widetilde{I}(t) = \sum_{s<t} I(s)q(s,t),\tag{2}$$

where $\widetilde{I}(t)$ is the value of intangible assets purchased on the time period s, and the multiplier $q(s,t)$, shows the part of the previous value of assets left at the moment t. Since a constant depreciation rate is used in business profit accountancy, the following equation will be used

$$q(s,t) = q^{(t-s)/4},\tag{3}$$

where the dimension of q is determined on the base of expert opinion. The depreciation rate $1-q$ shows how much the purchased assets depreciates on average in one year. Since equipment depreciates faster than buildings (constructions), it is might be expedient to divide the investment into two components: i) investment into implements of production (equipment and vehicles) and ii) other material investment. Under the assumption that only the first component of investment depreciates considerably, the whole investment volume $I(s)$ in equation 2 could be replaced by the first component mentioned correcting thereby the value of q.

4.2   Types of models

In this work, linear regression models and error correction models have been used. Let $Z(t)$ denote a vector of chosen regressors at the moment t. The simplest linear regression model is of the following shape

$$B(t) = \theta^T Z(t) + \varepsilon(t), \quad t = 1, \dots, n.\tag{4}$$

Here θ is the vector of model coefficients, $\varepsilon(t)$ is a stochastic error uncorrelated with $Z(t)$. Since during the resent years the economy of our country has been growing fast, the

variables of model (4) are not stationary. Therefore a model modification that describes relations of relative economic indicators is noteworthy:

$$b(t) = \theta^T z(t) + \varepsilon(t), \tag{5}$$

where the lower case letter denotes the ratio of the corresponding indicator and GDP, $b(t) = B(t)/Y(t)$. Here and in the sequel, coefficients denoted by the same letters can obtain different values. Equation 5 can be expanded including casual supplementary factors and dummy variables for seasonality and structural changes (change-points).

In the resent years the vector error correction models have been evermore widely used to describe the growing economy. They have also been successfully applied in modelling the Lithuanian macroeconomy (see in [13]). The error correction model is used when the time series analysed are nonstationary but they are cointegrated, i.e. their linear function is stationary. Models of this type are usually applied to logarithmic indicators, since the ratios of economic indicators could be characterised as more stable indicators. However the profit indicator could be negative in a discrete time period (quarter). Therefore by B^* we denote here a transformation

$$B^*(t) = \begin{cases} \log B(t), & \text{as } B(t) \geq 1, \\ 0, & \text{as } B(t) < 1, \end{cases}$$

$B^*(t)$ is said to be cointegrated with the stochastic $Z^*(t)$, if there exists a coefficient vector β such that

$$B^*(t) = \beta^T Z^*(t) + u(t), \tag{6}$$

where $u(t)$ is a stationary (in a wide sense) process.

Equation (6) describes long term relations of B^* and Z^*. Variable $u(t)$ is treated as a measure of deviation from the long term equilibrium $B^*(t) = \beta^T Z^*(t)$. The error correction model describing short term fluctuations is constructed on the base of the cointegration relation. The general form of the error correction model is

$$\Delta B^*(t) = \alpha u(t-1) + A(L)\Delta B^*(t) + \gamma^T \Delta Z^*(t) + D(t) + \varepsilon(t). \tag{7}$$

Here L denotes the lag operator: $L^k Z(t) = Z(t-k)$; $A(z)$ is a polynomial: $A(z) = a_1 z + \ldots + a_p z^P$; $D(t)$ is a deterministic component, which includes deterministic time trend, seasonal index, etc.; the vectors γ and β and coefficients $\alpha, \alpha_1, \ldots \alpha_p$ are unknown parameters of the model; $\varepsilon(t)$ denotes a stochastic error, which does not correlate with other components on the right side of equation (7). In addition, $\varepsilon(t)$ is usually assumed as a white noise term. The component $A(L)\Delta B^*(t) = \sum_{j=1}^{P} a_j \Delta B^*(t-j)$ is an autoregressive part of the model and the term $\gamma^T \Delta Z^*(t)$ shows the impact of a change in regressors.

The first component on the right side of equation (7) reflects the reaction of the dependent variable $B^*(t)$ to the deviation $u(t-1)$ in order to return to the long-term equilibrium. Therefore it is assumed that $\alpha < 0$. This condition can be explained even

without using the equilibrium terminology. Let us denote $\Phi(z) = (1-z)(1-A(z)) - \alpha z$. Then, equation (7) yields the equality

$$\Phi(L)B^*(t) = -\alpha\beta^T Z^*(t-1) + \gamma^T \Delta Z^*(t) + D(t) + \varepsilon(t). \tag{8}$$

Here we have got a model of conditional autoregression, where the usual assumption

$$\Phi(z) \neq 0, \quad \text{when} \quad |z| \leq 1. \tag{9}$$

assures the single set of parameter values. Since $\Phi(0) = 1$ and $\alpha = -\Phi(1)$ from equation (9) we obtain the inequality $\alpha < 0$.

4.3 The statistical evaluation of parameters

In the applied research, where linear regression schemes are used, the least squares parameter estimation method (LSM) is most common. To illustrate this method, we present the statistical estimation of model (4) parameter

$$\widehat{\theta} = \arg\min_\theta \sum_t \left(B(t) - \theta^T Z(t)\right)^2 \cdot w_t, \tag{10}$$

where w_t are chosen weights. In a particular case when $w_t \equiv 1$ the classical estimation of LSM is obtained. In econometrical theory it is recommended to use weights in the case, where residual $\varepsilon(t)$ is heteroskedastic, i.e. where its dispersion is not constant. Nevertheless there is a practical aspect of choice of the weights w_t. The applied models are only an approximation of the reality. The relations between B and vector Z of regressors are evolving, i.e., in the real economy even the parameters $\theta = \theta(t)$. Since the target of the model is to predict profit as accurate as possible, it is reasonable to seek more accurate approximation in the near past while the errors in further past are not that important. According to this logic the weight function is supposed to be growing even if residual $\varepsilon(t)$ of the model is homoskedastic.

Along with the LSM, the median estimation is also used in this work. In the latter case, the estimates of model parameters are defined by the equality

$$\widehat{\theta} = \arg\min_\theta \sum_t \left|B(t) - \theta^T Z(t)\right|. \tag{11}$$

The median estimations are more robust than statistics of type (10). This is important in the Lithuanian case because our economy experienced strong external impulses in the analysed period (Russian crisis, integration into EU, etc.). However, the lack of data prevent from the usage of outlier elimination methods. Even though the median estimates are less sensitive (i.e. robust) to data outliers in comparison with the LSM estimates, there arise a problem of calculating these estimates in practice, because the statistical programs for regression analysis usually have standard procedures only for the weighted LSM estimation. One way to avoid special programming of the median estimation procedure is a recursive calculation of the estimate using the weighted LSM procedure

$$\theta^{(k)} = \arg\min_\theta \sum_t \left|B(t) - \theta^T Z(t)\right|^2 \cdot w^{(k)}(t), \tag{12}$$

where $w^{(k)} = |B(t) - \theta^{T(k-1)}Z(t)|^{-1}$ are weights used in the k-th step. The recursive procedure starts from choosing the weights $w^{(1)}(t) \equiv 1$. In [14, 15], it is proved that the sequence of statistics $\theta(k)$ converges to the median estimate rather fast in the classical linear regression case.

In practice, if there are not many data, the aim is to construct much simpler models. Therefore, determination of significance level of factors becomes of the main importance. Determination of critical significance level (p-value) of the corresponding coefficient when verifying hypothesis of zero value of the coefficient is used mostly. Statistical programs calculate the p-value using the Student statistic based on the assumption of Gaussian distribution of white noise residual. However, in reality this assumption is not usually valid. Therefore the crossvalidation method is additionally used in this work. The model mean error estimates denoted by $\overline{\delta}$ are obtained by this method. Two models are compared when judging on the significance of a certain factor: in the first model, the factor is included into the regression vector and not in the second one. If in the latter case, the residual estimate $\overline{\delta}$ is not greater than that of the first model, then a conclusion on factor insignificance is drawn.

A special crossvalidation procedure called here *one leave out* or *jack-knife* has been used to calculate the average residual estimate $\overline{\delta}$ of the model. Let us discuss the case of model (5) for illustration. Let $\widehat{\theta}(\tau)$ denote the estimate of parameter θ obtained after eliminating the observation from the sample the time moment τ. The statistic $\overline{\delta}$ is defined by the equality

$$\overline{\delta} = \sum_{\tau=1}^{n} |\delta(\tau)|c(\tau), \tag{13}$$

where $\delta(\tau) = B(\tau) - \widehat{\theta}^T(\tau)Z(\tau)$, $c(\tau)$ are selected weights which sum is equal to 1.

4.4   Results of the modelling

Several models of profit indicators which could be used for profit prediction are presented below. Since the Lithuanian economy is still in transition and disposable series of observations are short, it is not reasonable to limit this set to only one model. When the number of observation increases, it will be possible to revise both the model parameters and the estimates of average residuals. In this work the statistics $\overline{\delta}$, and errors of profit prediction for 2005, obtained by the crossvalidation method (using the parameter estimates calculated by the observations until the end of 2004), have been used to characterise the accuracy of models. The values of these characteristics of accuracy, calculated for the models analyzed bellow are presented in the table at the end of the section.

On the basis of the abovementioned approximate formula of operating profit calculation on the micro level, first of all the regression equation has been identified

$$B(t) = \theta_1 Y(t) - \theta_2 F(t) - \theta_3 I(t) + \varepsilon(t). \tag{14}$$

All the three regressors were significant and the estimates of parameters met the a prior

assumption

$$\theta_i > 0, \quad i = \overline{1,3}. \tag{15}$$

As expected, the ADF test of the unit root showed nonstationarity of errors. The model had very bad accuracy statistics. The additional regressor $X(t)$ reflecting export influence poorly improved the modelling accuracy. In order to highlight export influence on profit, a several economy sectors such as manufacturing and transport were excluded and grouped into one sector. We call this group of sectors the production of which makes the major part of country export the first sector and the group of other sectors is called the second sector. $B_i(t)$ denotes the profit of the i-th sector. We also use this notation to denote other indicators. Thus, the economy profit is

$$B(t) = B_1(t) + B_2(t).$$

As shown by the analysis made, the profit forecast calculated with the aid of this decomposition is rather accurate. However, good results were obtained when indicators $B_1(t)$ and $B_2(t)$ were modelled in different ways. The profit indicator $B_1(t)$ was rather well described by the regression

$$B_1(t) = \theta_1 Y_1(t) - \theta_2 F_1(t) - \theta_3 \widetilde{I}(t-3) + \theta_4 X(t-3) + \varepsilon(t) \tag{16}$$

while in the case of the second sector this model gave poor results. The forecast accuracy of both the first sector profit and the whole economy profit has been increased (improved) when we moved to the ratio indicators

$$b(t) = \theta_1 - \theta_2 f(t) - \theta_3 i(t) + \varepsilon(t). \tag{17}$$

Here as above, the ratio between the indicator and GDP is denoted by the lower case letter. Equation (17) was expanded by including the bank credit average interest rate $r(t)$ and a dummy variable

$$d_1(t) = \begin{cases} 1, & \text{as } t_1 \leq t \leq t_2, \\ 0, & \text{in other cases.} \end{cases}$$

Here the time moment t_1 is the beginning of 1999 and t_2 is the end of 2000. The impact of the Russian crisis on Lithuanian economy was highest during these years. So the form of the expanded model is

$$b(t) = \theta_1 + \theta_2 f(t) + \theta_3 i(t) + \theta_4 r(t-4) + \theta_5 d_1(t) + \varepsilon(t). \tag{18}$$

Good results have also been obtained using equations of (18) type for the mentioned sectors separately. In the case of the first sector, the additional regressor $\widetilde{x}(t)$ (it is defined in the similar way as $\widetilde{X}(t)$ in equation (1) only replacing the absolute export difference indicator by the relative one $\Delta X(t)/X(t)$) was included

$$b_1(t) = \theta_1 + \theta_2 f_1(t) + \theta_3 \widetilde{i}_1(t-3) + \theta_4 d_2(t) + \theta_5 \widetilde{x}(t) + \varepsilon(t), \tag{19}$$

where the dummy variable

$$d_2(t) = \begin{cases} 1, & \text{as } t \geq t_3, \\ 0, & \text{in other cases.} \end{cases}$$

Here t_3 denotes the beginning of 2004 when the foreign trade conditions changed. The credit rate variable $r(t)$ and seasonal dummy variables included into equation (19) were not significant.

Applying the cointegration analysis, the error correction equation

$$\Delta B^*(t) = \alpha \big[B^*(t-1) + \beta_0 + \beta_1 Y^*(t-3) + \beta_2 \widetilde{I}^*(t-4) \big] + \theta_1 \Delta Y^*(t-3) + \varepsilon(t) \quad (20)$$

acquired not so bad accuracy properties. Here and hereafter, cointegration relation is indicated in the square brackets.

The indicators of the individual sectors cointegrate worse than that of the whole economy. Therefore modelling of the first sector profit indicators has not justified by this type of model, whereas the error correction equation of the second sector profit

$$\Delta B_2^*(t) = \alpha \big[B_2^*(t-1) + \beta_0 + \beta_1 Y_2^*(t-3) + \beta_2 \widetilde{I}_2^*(t-2) \big] + \theta_1 \Delta Y_2^*(t) + \varepsilon(t) \quad (21)$$

can be used for prediction of variable B_2.

The profit equations selected for profit prediction are presented below with parameter estimates under which their significance level (p-value), calculated in the standard way based on the Student statistics, are given in the brackets. The appropriate profit prediction errors $\overline{\delta}$, $\overline{\delta}^*$ are presented in Table 1. The statistics $\overline{\delta}$ is defined by equation (13) with weights $c(\tau) = 1/12$ if τ is for the quarter from the period of 2003–2005 or $c(\tau) = 0$ in other cases. The statistics $\overline{\delta}^*$ is appointed to estimate prediction errors not of the quarterly but yearly profit indicator. It is defined by equality

$$\overline{\delta}^* = \frac{1}{3} \big(|\delta(J_1)| + |\delta(J_2)| + |\delta(J_3)| \big),$$

where $\delta(J) = \sum_{\tau \in J} \delta(\tau)$. Here J_1, J_2, J_3 stands for one of the respective years 2003, 2004 and 2005. δ_{2005} denotes the percentage errors of profit forecast for 2005, calculated by the models that were estimated using data until 2004. The statistics $\overline{\delta}, \overline{\delta}^*$, are also given in Table 1 as a percentages of the corresponding sums $\sum_{\tau} B(\tau)c(\tau)$ and $\sum_{\tau=1}^{3} B(J_i)$.

Table 1. Accuracy characteristics of the profit models

Eq. No.	Profit tax base	δ_{2005}	$\overline{\delta}$	$\overline{\delta}^*$
(22)	OP	2.6	7.8	5.9
(23)	PBT	2.7	5.4	3.6
(24)	PBT$^+$	2.2	5.4	3.6
(25), (26)	OP	0.9	6.3	1.5

4.5 Estimated models

$$
\begin{aligned}
b(t) = &-0.177 + \underset{(<0.001)}{0.928} \cdot f(t) - \underset{(<0.001)}{0.624} \cdot \widetilde{i}(t) - \underset{(<0.001)}{8.304 \cdot 10^{-3}} \cdot r(t-4) \\
&+ \underset{(0.002)}{0.094} \cdot \widetilde{x}(t) - \underset{(0.005)}{0.029} \cdot d_1(t) + \varepsilon(t).
\end{aligned}
\tag{22}
$$

Here the profit indicator is OP, and $\widetilde{i}(t)$ is calculated using the indicator of the material investment into equipment and vehicles. The model parameters are estimated using the median method with alternating weights $w^{(k)}(t)$ defined for this method.

$$
\begin{aligned}
b(t) = &-0.139 + \underset{(<0.001)}{1.150} \cdot f(t) - \underset{(0.002)}{0.659} \cdot \widetilde{i}(t) - \underset{(<0.001)}{9.684 \cdot 10^{-3}} \cdot r(t-4) \\
&+ \underset{(0.085)}{0.071} \cdot \widetilde{x}(t) - \underset{(0.007)}{0.042} \cdot d_1(t) + \varepsilon(t).
\end{aligned}
\tag{23}
$$

Here the profit indicator is PBT, and $\widetilde{i}(t)$ is calculated using the indicator of the material investment into equipment and vehicles. The model is estimated using LS method $(w(t) = 1)$.

$$
\begin{aligned}
\Delta B^*(t) = &-0.869 \big[B^*(t-1) + \underset{(<0.001)}{10.900} - \underset{(<0.001)}{1.840} \cdot Y^*(t-3) - \underset{(0.090)}{0.639} \cdot \widetilde{I}^*(t-4) \big] \\
&- \underset{(0.001)}{0.942} \cdot \Delta Y^*(t-3) + \varepsilon(t).
\end{aligned}
\tag{24}
$$

Here the profit indicator is PBT^+ and $\widetilde{I}(t)$ is calculated using only a part of material investment, i.e. the material investment into equipment and vehicles. The model is estimated using LS method $(w(t) = 1)$.

$$
\begin{aligned}
b_1(t) = &-0.120 + \underset{(0.004)}{2.302} \cdot f_1(t) - \underset{(<0.001)}{0.591} \cdot \widetilde{i}_1(t-3) \\
&- \underset{(0.041)}{0.042} \cdot \widetilde{x}(t) + \underset{(0.003)}{0.013} \cdot d_2(t) + \varepsilon(t).
\end{aligned}
\tag{25}
$$

Here the profit indicator is OP of the first sector. The model is estimated using LS method $(w(t) = 1)$.

$$
\begin{aligned}
\Delta B_2^*(t) = &-0.608 \big[B_2^*(t-1) + \underset{(<0.001)}{23.198} - \underset{(0.009)}{2.148} \cdot Y_2^*(t-3) - \underset{(0.015)}{1.263} \cdot \widetilde{I}^*(t-2) \big] \\
&+ \underset{(0.078)}{0.586} \cdot \Delta Y_2^*(t) + \varepsilon(t).
\end{aligned}
\tag{26}
$$

Here the profit indicator is OP of the second sector and $\widetilde{I}(t)$ is calculated using the indicator of the material investment into equipment and vehicles. The model is estimated using LS method $(w(t) = 1)$.

5 Regression models of the profit tax revenue

As stated in Section 2, approximation of the profit tax revenue $T(t)$ by the product $B(t)R(t)$, where $B(t)$ denotes the tax base and $R(t)$ – the tax rate, is too rough (general and inaccurate) because of advanced payments of the profit tax of enterprises and transfer of the losses incurred in the previous years to the current year. Nevertheless, it is doubtless that the mentioned product is a significant regressor for the variable $T(t)$. However, it is not clear a priori which of the profit indicators suits for the role of profit tax base best: the operating profit (OP) of nonfinance enterprises, their profit before tax (PBT), or PBT of the enterprises which worked profitably during the period t. The statistical analysis made by the authors has shown that use of the latter indicator as the tax base yields lower accuracy of regression in comparison with the other two indicators. Therefore $B(t)$ will denote PBT or OP of nonfinance enterprises in the sequel. According to the regulations of advanced payments of the profit tax, we will define another important regressor denoted by $B_0(t)$. There is a due date defined by the law until which enterprises are obliged to submit their financial account of the last year to the STI. Let $J(t)$ denote the last year of the previous years when the financial account was supposed to be submitted until the time moment t. The value of advanced payment attributed to the enterprise at the quarter t is defined by its earned profit in the year $J(t)$. Therefore the variable $B_0(t)$ is defined by the equality

$$B_0(t) = \frac{1}{4} \sum_{\tau \in J(t)} B(t). \tag{27}$$

A certain seasonality typical to the revenue $T(t)$ has been exposed since 2003 when the new amendments of the profit tax law concerning due dates of the tax payment came into force. Therefore, there are two significant dummy variables in the cointegration equation of logarithmic indicators

$$T^*(t) = \beta_0 + \beta_1 B^*(t) + \beta_2 B_0^*(t) + \beta_3 D_3(t) + \beta_4 D_4(t) + u(t). \tag{28}$$

Here $D_i(t)$ is equal to 1 if t corresponds to the i-th quarter of the year starting from 2003, and it is equal to 0, in other cases; $i = 3, 4$; β_i denote coefficients after estimating of which by the LS method, the ADF test showed the stationarity of residuals $u(t)$. The regressor $R^*(t)$ was not significant statistically. In the error correction equation presented below the significance level (p-value) of coefficients is written under the each of them in the brackets as in Section 3

$$T^*(t) = \underset{(<0.001)}{-0.729} \big[T^*(t-1) - \underset{(<0.001)}{0.667} \cdot B^*(t-1) - \underset{(<0.001)}{1.849} \cdot D_3(t) - \underset{(0.002)}{0.809} \cdot D_4(t) \big]$$
$$+ \underset{(0.017)}{0.477} \cdot \Delta B^*(t-1) + \underset{(0.025)}{0.761} \cdot \Delta B_0^*(t-1) + \varepsilon(t). \tag{29}$$

The parameters of model (29) were estimated by the weighted least squares method with linearly growing weights $w(t) = 1 - \frac{t}{n}$, $t = 1, \ldots, n$. The average error $\overline{\delta}$ is

defined similarly as in Section 3 was 8.0 % for the quarter values of revenue and 7.2 % for the yearly values of revenue (in the period of 2003–2005). The error of profit forecast for 2005 obtained by the crossvalidation method was 6.4 %. It ought to be stated that the accuracy of this model is not high and the main reason is rather a large number of statistically estimated parameters in view of the short observation series. However, in the future a significant improvement of the accuracy could be expected in case we have more observation data and revise the coefficients of equation (29) accordingly.

At the present-day situation, when forecasting revenue the better results have been obtained using the linear regression model for nonlogarithmic indicators. In order to get stationary errors, the indicators of ratio between the tax or the profit and GDP have been analysed. After elimination of insignificant indicators the following equation has been obtained

$$
\frac{T(t)}{Y(t)} = 6.308 \cdot \underset{(<0.001)}{\frac{B(t)}{Y(t)}} \cdot R(t) + 3.789 \cdot \underset{(0.022)}{\frac{B_0(t)}{Y(t)}} \cdot R(t) - 2.937 \cdot \underset{(0.004)}{\frac{L(t-3)}{Y(t-3)}} \cdot R(t)
$$
$$
+ 0.017 \cdot \underset{(<0.001)}{D_3(t)} + 0.013 \cdot \underset{(<0.001)}{D_4(t)} + \varepsilon(t). \tag{30}
$$

Here $L(t)$ denotes the losses of profitless enterprises during the period t. The model coefficients have been estimated in the same way as that of equation (29). Analysing the accuracy of the tax revenue prediction by model (30), the following results have been obtained: $\overline{\delta} = 5.1$ % for the quarterly forecast, $\overline{\delta} = 6.0$ % for the yearly forecast, and the error of forecast for 2005 was 3.6 %. For both (29) and (30) models, the prediction accuracy worsened if the median method was used to estimate the parameters. As mentioned above, the latter method is useful when outliers are presented in data. However, when analysing the time series of the ratio indicator $T(t)/B(t)$, outliers have not been observed.

In what follows, we estimate what prediction errors for the profit tax revenue are obtained when the values of profit indicators used in equation (29) and (30) are not known but are predicted using models described in the Section 3. To estimate the accuracy, the same characteristics are used as in Table 1, where the set of selected equations is considered as a compound model of $T(t)$. The compound prediction results are presented in Table 2.

Table 2. Accuracy characteristics of compound prediction

Eq. No.	Profit tax base	δ_{2005}	$\overline{\delta}$	$\overline{\delta}^*$
(29) (22)	OP	6.7	8.0	7.2
(30) (23)	OP	6.0	5.1	6.0
(29) (24) (26)	OP	5.1	7.9	6.8
(30) (25) (26)	OP	3.7	6.6	5.8

After reviewing the presented results, it could be stated that the most recommended procedure for profit tax revenue prediction at present includes equations (25), (26) and

(30); moreover it is reasonable to use the OP of nonfinance enterprises as the tax base. Using this algorithm the yearly prediction error should not exceed 6 % disregarding the influence of the errors of GDP and fund of wages forecast. However the effect might be not so significant since Ministry of Finances predicts them rather precisely.

References

1. E. Mačiulaitytė, The ways of profit tax modeling in Lithuania, *Ekonomika*, **73**, pp. 57–67, 2006.

2. S. C. Kumbhakar, Estimation of Profit Function When Profit Is Not Maximum, *American Journal of Agricultural Economics*, **83**(1), pp. 1–19, 2001.

3. J. Bradley, A. Kangur, I. Kearney, *HERMIN HE4. A medium-term macro-sectoral model of Estonia: structure, properties and forecasts*, presented at Ministry of Finance Seminar, Tallinn, 2001.

4. D. Rea NBNZ-DEMONZ: A dynamic equilibrium model of New Zealand, *Economic Modelling*, **13**, pp. 91–165, 1996.

5. A. Willman, M. Kortelainen, H.-L. Mannisto, M. Tujula, *The BOF5 Macroeconomic Model of Finland. Structure and Equations*, Bank of Finland discussion papers, No. 10, 1998.

6. A. Willman, M. Kortelainen, H.-L. Mannisto, M. Tujula, The BOF5 macroeconomic model of Finland, structure and dynamic microfoundations, *Economic Modelling*, **17**, pp. 275–303, 2000.

7. D. Celov, F. M. Andersen, E. Vilkas, D. A. Grinderslev, Macro-Econometric Model for Lithuania LITMOD, `http://www.ekm.lt/lt/strategija/doc/litmod2.pdf`, 2003.

8. P. Van den Noord, *The Size and Role of Automatic Fiscal Stabilizers in the 1990s and Beyond*, OECD Economics Department Working Papers, No. 230, 2000.

9. S. Kennedy, N. Luu, S. Morling, L. Yeaman, Fiscal Policy in Australia: Discretionary Policy and Automatic Stabilisers, in: *ANU Macroeconomic Conference*, 2004.

10. Y. Hsing, Estimating the Laffer Curve and Policy Implications, *Journal of Socio-Economics*, **25**(3), pp. 395–401, 1996.

11. A. Budrytė, E. Mačiulaitytė, Biudžeto pajamų iš pelno mokesčio prognozavimas, *Pinigų studijos*, **3**, pp. 5–27, 2005.

12. R. Rudzkis, V. Kvedaras, Lietuvos eksporto tendencijos ir eknometriniai modeliai, *Pinigų studijos*, **4**, pp. 13–35, 2003.

13. R. Rudzkis, V. Kvedaras, A Small Macroeconometric Model of the Lithuanian Economy, *Australian Journal of Statistics*, **34**(2), pp. 185–197, 2005.

14. R. H. Byrd, D. A. Pyne, Convergence of the iteratively reweighted least squares algorithm for robust regression, *Technical report*, **313**, Dept Math. Sci., John Hopkins Univ., Baltimore, MD, 1979.

15. R. Yarlagadda, J.B. Bednar, T.L. Watt, Fast algorithms for l_p deconvolution, *IEEE Trans. Acoust., Speech, Signal Processing*, **43**, pp. 174–182, 1995.

16. I. Vetlov, Baltijos šalių ekonomikos augimo apskaita, *Pinigų studijos*, **3**, pp. 14–34, 2003.

17. J. Glen, K. Lee, A. Singh, Corporate profitability and the dynamics of competition in emerging markets: a time series analysis, *The Economic Journal*, **113**, pp. 465–484, 2003.