Research of Estimates of Tax Revenue: An Overview[#]

Ondřej BAYER^{*}

Introduction

Estimating tax revenue is a crucial for the creation of state budgets, because each government is able to determine exactly just spending, implementation of the revenue budget is only able to estimate. Given that tax revenues are mostly 95% of all government revenues are accurate tax assessments essential to the quality of fiscal policy.

The aim of this paper is to describe the procedures used for estimating tax revenues in the Czech Republic and abroad.

Many scientific papers deal with problems of tax revenue estimating, but some of them only marginally. The author based on the best possible publicly available publications.

In the case of the Czech Republic devoted to the topic Klazar (2003), who in his work creates estimation models for the Czech tax system and subsequently test the accuracy of models with the Ministry of Finance. This work is subsequently extended by Bayer (2011a) and Podhradská (2008), who extend these models. Bayer (2011b) devotes his work creating micro-simulation with an emphasis on quality testing of government predictions. Špalek – Moravanský (2005) also test the bias of government tax estimates and deeply concerned with errors estimates. Bayer (2012b) using a micro-simulation tests the possible impacts of unification of VAT rates to Czech taxpayers.

The essential foreign publication is Jenkins et al. (2010), which comprehensively describes the problem of tax revenue forecasting and tax elasticities using microsimulation. For micro-approach are used data from the USA. Ram (1991) also focuses on estimating the tax elasticities.

[#] This article has been elaborated as one of the outcomes of research project *The Effect* of *Tax and Expenditure Instruments on Macroeconomic and Microeconomic Efficiency* which id granted by Internal Grant Agency of University of Economics in Prague under registered number F1/30/2010.

^{*} Ing. Ondřej Bayer – student Ph.D.; Department of Public Finance, Faculty of Finance and Accounting, University of Economics, Prague, W. Churchill Sq .4, 130 67 Prague, Czech Republic;<xbayo00@vse.cz>.

Johnson – Lambert (1989) tested the relationship between tax revenue and GDP growth, this work has been freely extended by Lee – Gordon (2005). Creedy – Gemmel (2006) evaluates all the factors affecting the growth of tax revenue.

Buettner – Kauder (2010) are based on the quality of estimates. The work itself is defined as a continuation of research projects assessing the quality of estimates in the individual federal states in the U.S. The authors are trying to evaluate the quality of predictions of the OECD countries. The study itself does not address the precise consequences of error estimation, but only evaluate the quality of quantitative forecasting of the countries. The variance analyzes using methodologies MSFE (mean square error of the estimate). SDFE (standard deviation error estimate) and RMSFE (root-mean forecasting error, simply put, is a combination of mean error of estimation and standard deviations). The article deals with the evaluation of the quality of estimates in relation to the quality of estimates of future economic development with GDP forecasts. Another aspect of the evaluation of the quality of estimates is an institutional arrangement estimates of when some studies suggest that the most effective is the use of state authorities not only directly but also private. Keene - Thompson (2007) have studied the variations in government forecasts of New Zealand. Auerbach (1999) tests the quality of the predictions of the U.S. with potential implications for their adjustment methodology refinement.

The actual contribution is divided into several parts, with the first part deals with the general distribution of predictions. The next two sections are devoted to different types of approach to modeling (namely GDP static models, regression analysis and time series analysis) Last part is dedicated to quality assessment predictions and the reasons for distortion. The research results are summarized in a brief conclusion.

Approaches to estimating

In the case of construction estimates, these estimates can be divided according to several well-known criteria. When tax estimates can be used as the first criterion for the division to use overall view of the tax system. Bayer (2011a) uses the distinction between qualitative and quantitative estimates.

Qualitative estimates are based on a more subjective basis, where the greatest emphasis is placed on personal experiences and beliefs of forecaster. The actual quality of these estimates is directly proportional to

experience forecaster, although it would seem that this method cannot be sufficiently precise, so in certain cases, the results are better than using econometric methods. The actual quantitative estimation permits automatic and subconscious balancing monitored data based on time series analysis. Perception of a subjective component in these estimates advantages rapid interpretation and evaluation of the consequences of sudden radical changes in the monitored variables¹.

In the case of estimating the tax revenue is possible high accuracy of quantitative estimates due representation of a relatively strong behavioral components of the explanatory variables. Can generally say that qualitative analyzes are quite useful when you need to create ad hoc analyzes or estimated value is based more on qualitative indicators.

Quantitative estimation is based on econometric and statistical methods. This methodology is much more widespread in the approach to estimating tax revenues, because it is better comparable and measurable than quantitative analysis, and therefore is used in most professional publications. Problem using a quantitative approach is the design concept of rational models. These models must meet the requirements for high accuracy of prediction, together with the simplest possible interpretations. In the case of tax revenues is quite difficult to cover all possible factors affecting tax revenues, but Klazar (2003) argues that most of these factors has only minimal explanatory value, or the influence of some model variables cancel each other out. Another limitation of the quantitative approach is a strong dependence on the quality of input data, which follows from the assumptions of various econometric and statistical approaches.

In the case of quantitative modeling methods are mainly used regression analysis and extrapolation of time series, or their derivatives, as shown in literature. Quantitative approach also faces the problem of the inclusion of some factors can be difficult behavioral economics. Sometimes it is necessary to assume that these factors have only a minimal effect, or take advantage of the econometric analysis and describe these phenomena as a random variable.

Another distribution of design estimates is a preview of the economic system as a whole. In the case of micro-economic approach is a method

¹ For example is the recent economic crisis, the sophisticated economic models were unable to predict it, which is caused by the fact that even self-accuracy time series need a long period to be possible to predict and respond to similar shocks.

called "Bottom-Up", where the total estimates are based on estimating the tax revenue from the smallest possible taxable units (mostly individual). Overall view of the reference system is then obtained by gradual aggregation of the individual components.

Microsimulation itself may have for modeling of tax revenue several interesting findings. Bayer (2011b) tests accuracy of government's predictions by using microsimulations, which can be used as well as the results for the interpretation of the time delay introducing legislative changes.

Jenkins et al. (2000) the greater part of his work focuses microeconomic approach, when used as a microsimulation of tax proposals to various solutions. This approach is also able to identify potential redistribution effects associated with tax arrangements, as evidenced for example Bayer (2012a, p.10).

In the case of macro-economic approach (Top-Down method) is the modeling of tax revenues in higher administrative level - usually government. This method is not so demanding on data resources and their processing and results could be functional econometric models based on regression analysis, for example Bayer (2012b).

Most of the literature favors the possibility of using Top-Down estimates example Klazar (2003) and Buettner – Kauder (2010).

Forecasting methods based on GDP

Can definitely be said that GDP, or its derivatives, is the most used variable for the design of tax estimates. Generally the construction of models based on GDP describe as the relationship between the tax bases, tax rates and the relevant tax return. As a premise can be used known close positive correlation between GDP and tax revenue. For example, in the case of personal income tax is likely that the growth of GDP increases of wages, and therefore will be higher tax revenue at a constant tax rate. Models using GDP can be static or dynamic.

In the case of dynamic modeling is the assumption that amends the tax base in real time according to the available data. Dynamic modeling assumes that in the case under consideration changes in the tax system will be subject to certain behavioral assumptions. Unfortunately model estimations of behavioral effects are not completely accurate (Jenkins et al., 2000). At the time of the introduction of changes enter to the operating model many discrete factors that may affect the reference tax base among the estimated values. For accurate dynamic models are needed very extensive, accurate and long time series, while resulting models themselves are very complicated, and therefore, most countries are not used.

Static versus dynamic models are relatively simple, since they are not so much dependent on the quality and volume of input data. These models assume time stable changes to the tax base. The disadvantage of this approach is the high degree of dependence between the quality of the model estimating tax revenues and future development in the adjustment of the tax base, here you can find some parallels in the measurement of tax revenues that are expressed as adjusted cash principle.

Static models can be according to Jenkins et al. (2000), divided into models with constant gain and proportional models (proportional adjustment). Both models aim to identify changes in tax revenue based on discrete components (legislative changes) or dependence on economic development.

In the case of models with a constant increase are expected future tax revenues forecasted through tax base multiplied by the rate of the last period of the current year. The result of this estimate implies the possible impacts of legislative changes. The disadvantage of this approach is always constant distribution of the tax mix, because it assumes only potential gradualist trend in all taxes based on the basis of economic growth. The construction of this model is quite difficult because you need to have high quality and detailed input data, most of it can be applied to the specific excise duty (Jenkins et al., 2000).

Proportional models are complicated, because they assume delay implementation of discrete variables. The output of this model is the relationship of tax revenue to economic activity, since it eliminates the effects of discrete variables. In general, this approach assumes relative stability of the tax system for the monitored period, but on the other hand, this requires only basic input data. The construction of this model can be according to Jenkins et al. (2000) described as follows:

$$AT_{n-1} = T_{n-1} \cdot [T_n / (T_n - D_n)], \tag{1}$$

where AT_{n-1} = adjusted tax revenues at n-1, T_{n-1} = tax revenues at n-1, T_n = tax revenues at *n*, D_n = revenues collected for discretionary changes at *n*.

The result of this model are estimates of what would be the tax revenues in year n-1, if n applied conditions of course, this algorithm can be used for a longer time delay.

This calculation can be according to Jenkins et. al. (2000) used for the calculation of elasticity:

$$E_{TY} = \frac{AT_2 - AT_1}{Y_2 - Y_1} \cdot \frac{Y_1}{AT_1}$$
(2)

where E_{TY} = tax elasticity, AT_n = adjusted tax revenues at n, Y_n = product at n,

The implementation of these formulas is derived based on introduction and subsequent estimating the correct functional variables for individual taxes. For example, wages, employment, inflation in the case of PIT. From this relationship has already possible to create different prediction models based on regression analysis, when explaining variable are adjusted tax returns and explanatory variables individual macroeconomic variables (Jenkins et al., 2000). In the case of using a logarithmic regression function:

$$\ln AT_j = \gamma + \delta \cdot \ln Y_j \tag{3}$$

where AT_j = adjusted tax revenue, γ = regression constant, δ = regression parameter.

 δ is actually tax elasticity:

$$\delta = \frac{\Delta AT_j / AT_j}{\Delta Y_j / Y_j} = \frac{\% AT_j}{\% Y_j} \tag{4}$$

where AT_j = adjusted tax revenue, Y = product.

And this construction elasticity has already used to estimate the potential future tax revenues. The actual procedure of static structure proportional models can be according to Jenkins et al. (2000) summarized:

- 1. Convert all categories of taxes from nominal to real terms.
- 2. Obtain and convert all corresponding nominal discretionary changes to real.
- 3. Calculate the adjusted coefficient and cumulative coefficient to reflex discretionary changes.
- 4. Multiply real tax revenues with cumulative coefficient of discretionary changes to get an adjusted tax series.
- 5. Regress the respective adjusted tax revenues with corresponding tax base to find the tax elasticity.
- 6. Use estimated tax elasticity to forecast the future revenue stream.

Econometric methods

The design of tax estimates is clearly most commonly used regression analysis. This statistical method examines the relationship between one explaining variable (tax revenue) and one or more explanatory variables (commonly used any macro-indicators). The result of this analysis is the regression function, which can take the following forms:

a) linear:

$$Y = b_0 + b_1 \cdot x, \tag{5}$$

- where Y = dependent variable, b_0 = regression constant, b_1 = regression parameter, x =independent variable.
- b) parabolic:

$$Y = b_0 + b_1 \cdot x + b_2 \cdot x^2,$$
 (6)

- where Y = dependent variable, b_0 = regression constant,

 - b_1 = regression parameter,
 - = independent variable, x
 - b_2 = root regression parameter.
- c) logarithm:

$$Y = b_0 + b_1 \cdot \ln x, \tag{7}$$

where Y = dependent variable,

 b_0 = regression constant, b_1 = regression parameter, x = independent variable.

To evaluate the quality of the regression analysis is crucial coefficient of determination (\mathbb{R}^2), which indicates how much variance observed variables are a model able to explain. Other criteria for assessing the quality of regression models are standard statistical tests (F-test, t-test), in the case of autocorrelation Durbin-Watson statistic. When using a multiple regression model refines adding more explanatory variables will be used in the partial coefficient of determination (\mathbb{R}^2 part.) and must be dealt with the issue of multicollinearity (Bayer, 2011a).

Another possible approach is to use time series analysis, where it is assumed that there is no better explanatory variable than the actual development of the monitored variables over time. There are two basic types of time series: additive and multiplicative. In the case of additive time series the resulting function is:

$$y_t = T_t + S_t + C_t + e_t,$$
 (8)

where y_t = dependent variable at t, T_t = trend component at t, S_t = seasonal component at t, C_t = cyclical component at t, e_t = random variable at t.

and multiplicative function is:

$$y_t = T_t \cdot S_t \cdot C_t \cdot e_t \tag{9}$$

where y_t = dependent variable at t, T_t = trend component at t, S_t = seasonal component at t, C_t = cyclical component at t, e_t = random variable at t.

T always expresses the trend component, which reflects the main tendency of development of fixed values of the analyzed indicators over time. The trend can be increasing or constant, when indicator values fluctuate around a essentially constant level. With seasonal component explains periodic deviations from the trend due to specific conditions during the observation period of less than one year. The cyclical component C describes specific recurring deviations observed in the time horizon longer than one year. The last component is a random variable e, which cannot be described by any of the functions of time. The most important of these components is a component of the trend (linear, parabolic or exponential) that explains and predicts most of the observed variables. Other ingredients only serve to refine the explanation of periodic and significant deviations from the trend (Bayer, 2011a).

Kubátová et al. (2012) is based on macroeconomic theory and try to connect approaches regression analysis and time series analysis, which derives the following relations:

$$T_{t} = Ty_{t} + Ta_{t}, \qquad (10)$$
where $T_{t} = \text{total tax revenue at } t,$
 $Ty_{t} = \text{income tax revenue at } t,$
 $Ta_{t} = \text{autonomous tax revenue at } t,$

Also he assumes that autonomous tax revenue is constant and depends only on the rate of autonomous taxes ta. Therefore applies

$$T_a = f(t_a) = const.,\tag{11}$$

or dynamic approach

$$T_{at} = f(t_{at}),\tag{12}$$

and if t_a constant, then

$$T_{at} = f(t), \tag{13}$$

In the case of income taxes, the resulting revenue is function of income Y and the tax rate of income taxes ty. Therefore he applies:

$$T_{y} = f(Y, t_{y}), \tag{14}$$

dynamic approach

$$T_{yt} = f(Y_{t,t}y_{t}) \tag{15}$$

In the case of abstracting from the tax rate in the selected data, replacing income other functional independent variable (B), and where the dynamic autonomic components can therefore he said:

$$Tx_t = Txb_t + Ta_t, \tag{16}$$

where
$$Tx_t = \tan revenue x \tan t t$$
,
 $Txb_t = \operatorname{part} of \tan revenue dependent on the B \operatorname{at} t$
 $(Txb_t = f(B_t))$,
 $Ta_t = \operatorname{autonomous part} of \tan revenue \operatorname{at} t (Ta_t = f(t))$.

If it can be assumed that every tax has an autonomous and dependent part and this ratio can be estimated using the parameters. It is possible to combine them into a single estimate as regression and time series analysis. Autonomous part is clearly identifiable by time analysis, while the dependent part of the tax function can be determined by using regression analysis, as described in the following equation, where they were replaced by functional relationships estimated coefficients (Kubátová, 2012).

$$Tx_t = b_0 + b_1 \cdot B_t + b_2 \cdot t, \tag{17}$$

Where
$$Tx_t = \text{tax revenue } x \text{ tax at } t$$
,
 $b_0 = \text{regression constant}$,
 $b_1 = \text{regression parameter dependent for parameter } B$,
 $b_2 = \text{regression parameter dependent for time variable}$.

Unfortunately, the results of this connection so far do not increase accuracy. Therefore this approach does not show higher accuracy than the actual regression analysis yet.

Accuracy of estimates and the reasons for bias

In the case of using macrosimulations is quite crucial to evaluate the quality of the resulting prediction models and attempt to explain any distortion. The criterion for the quality prediction model can be expressed not only its accuracy such as coefficient of determination, but also the demands of the model to the input data.

Publication devoted to design their own specific regression models (Bayer, 2011a; Podhradská, 2008; Klazar, 2003) mostly use the possibility of ex-post analysis, where they use of feasibility of reducing the time-series regression of a certain period and then generate a forecast for a known period. Alternatively, they compare their estimates with estimates of the Ministry of Finance in order to determine whether constructed models are more accurate than the official models.

Another possible approach to quality assessment tax estimates is a group of publications focusing on the quality of government predictions with a solution proposal to refine these predictions based on the experiences of other states. In this case, it is not only income tax, but the total budget revenues – in most of countries, taxes are 99 % of the budget, so this kind of research can be considered a tax estimates.

In general we can say that the quality of government predictions is a crucial institutional arrangement sites responsible for the estimates. Buettner – Kauder (2010) argues that the best option is a "competition" between public and private institutions, when are much less demonstrable inaccuracies.

The actual methodology of measuring the quality of the predictions are mostly based on statistical indicators for measuring predictions the most often applied indicators are:

$$MAE = \frac{\sum_{t=1}^{N} |E_t|}{N}, \qquad (18)$$
where MAE = mean absolute error,
 E_t = mean at $t,$
 N = number of observations.

$$MAPE = \frac{\sum_{t=1}^{N} \left|\frac{E_t}{Y_t}\right|}{N}, \qquad (19)$$
where $MAPE$ = mean percentage absolute error,
 E_t = mean at $t,$
 Y_t = portion at $t,$
 N = number of observations.

$$MSE = \frac{\sum_{t=1}^{N} |E_t^2|}{N}, \qquad (20)$$
where MSE = mean squared error,
 E_t = mean at $t,$
 N = number of observations.

$$RMSE = \sqrt{\frac{\sum_{t=1}^{N} |E_t^2|}{N}}, \qquad (21)$$
where $RMSE$ = Root mean squared error,
 E_t = mean at $t,$
 N = number of observations.

In the evaluation of the errors is vital distribution of statistical errors. Errors in forecasts can be divided into political, economic, and technical errors.

Political error is uncertainty estimation caused by an unexpected legislative change. The forecast is based on a known state of legislative amendments, or allows for planned modifications. In the legislative process, it is possible that the expected changes are adjusted or if they arise unexpected new legislation changes. The impact of these changes cannot be included into the already processed prediction (Bayer, 2011a).

As noticeable errors forecasting appears economical errors. These are typical errors that arise in the preparatory stage of prediction, which it is generated by the macroeconomic framework (Špalek – Moravanský, 2005). Most of these errors are in determining the future development of the economy or sudden economic changes.

Other factors which cannot be included in the previous groups are referred to as technical errors. According Špalek – Moravanský (2005) these causes can be divided into two groups:

"Errors in predicting aggregate elasticity of tax revenues to their base (for example, errors associated with sudden shifts in the structure of demand, output or income in connection with a widely distributed effective tax rates), these sudden changes of aggregate elasticities are often associated with the process of economic cycles. Errors in predicting income elasticities of individual taxes to the appropriate tax base (i.e. elasticities applied to a set of economic factors). These errors occur in cases where the tax system leaves a relatively large field effect behavior change in economic indicators (tax evasion, exploit loopholes in the law or tax fraud)."

Identification of errors is very important, because with a sufficiently large number of observations can be adjusted to a more precise methodology.

The most serious error in terms of prognostics is an economic error. Economic error means error in estimating the future development of the economy. To eliminate this error, the introduction of such uses selected macroeconomic indicators, which is the best statistically significant explanatory variable for the tax return (Bayer, 2011a).

Conclusion

Aim of this paper was to approach most commonly used methods and indicators for tax estimates. Generally, it can be argued that this objective was achieved, but of course each publication regarding tax forecasts have small differences in calculation methodology. For descriptions of the actual design of each model is quite difficult to find relevant methods from the literature in addition to Czech. On the other hand, foreign publications deal mainly with quality assessments forecasts or using microeconomic approach. The results of this field research reveal that this area is covered with relatively little research work that is publicly available (Ministry of Finance do not reveal their models). The paper shows that if it is necessary to create tax estimates of the tax changes, together with their impact is much better use microeconomic approach, while for long-term analyzes are more favorable estimation methods based on the Top-Down approach.

References:

- Auerbach, A. J. (1999): On the Performance and Use of Government Revenue Forecasts. [on-line], Berkeley, University of California, c1999, [cited 8th August, 2013], http://escholarship.org/uc/item/8h845262>.
- [2] Bayer, O. (2011a): *Měření a odhadování daňových příjmů*. Praha, VŠE v Praze, 2011. Diploma thesis.
- [3] Bayer, O. (2011b): Vliv daňové predikce: ex ante odhady a ex post hodnocení přesnosti v České republice. Český finanční a účetní časopis, 2011, vol. 6, no. 1, pp. 42-54.
- [4] Bayer, O. (2012a): Development of Social Insurance Revenues in the Czech Republic. In Špalková, D. (ed.): Modern and Current Trends in the Public Sector Research. Brno, Masarykova univerzita, 2012.
- [5] Bayer, O. (2012b): *Reforma sazeb daně z přidané hodnoty v České republice*. Český finanční a účetní časopis, 2012, vol. 7, no. 1, pp. 82-91.
- [6] Buettner, T. Kauder, B. (2010): Revenue Forecasting Practices: Differences across Countries and Consequences for Forecasting Performance. [on-line], Fiscal Studies, 2010, vol. 31, no. 3, pp. 313-340, [cited 8th August, 2013],
 http://www.sbs.ox.ac.uk/centres/tax/symposia/Documents/2010/Bu

<http://www.sbs.ox.ac.uk/centres/tax/symposia/Documents/2010/Bu ettner%20Revenue_Forecasting_April_2010.pdf>.

- [7] Creedy, J. Gemmell, N. (2006): *Modelling Tax Revenue Growth*. [on-line], Cheltenham, Edward Elgar Publishing, c2006, [cited 8th August, 2013],
 .
- [8] Jenkins, G. P. Kuo, C. Y. Shukla, G. (2000): Tax Analysis and Revenue Forecasting: Issues and Techniques. [on-line], Boston, Harvard Institute for International Development, c2000, [cited 8th August, 2013],

<http://jdintl.econ.queensu.ca/publications/qed_dp_169.pdf>.

- [9] Johnson, P. Lambert, P. (1989): Measuring the Responsiveness of Income Tax Revenue to Income Growth: A Review and Some UK Values. Fiscal Studies, 1989, vol. 10, no. 4, pp. 1-18.
- [10] Keene, M. Thomson, P. (2007): An Analysis of Tax Revenue Forecast Errors. [on-line], Wellington, New Zealand Treasury, 2007, [cited 8th August, 2013], <http://treasury.govt.nz/publications/research-policy/wp/2007/07-02/twp07-02-full.pdf>.
- [11] Klazar, S. (2003): *Efektivnost predikcí daňových příjmů v ČR*. Praha, University of Economics, Prague, 2003. Doctoral thesis.
- [12] Kubátová, K. et al. (2012): Vliv daňových a výdajových nástrojů na mikroekonomickou a makroekonomickou efektivnost. Praha, Wolters Kluwer, 2012.
- [13] Lee, Y. Gordon, R. H. (2005): *Tax structure and economic growth*.
 [on-line], Journal of Public Economics, 2005, vol. 89, no. 5, pp. 1027-1043, [cited 8th August, 2013],
 http://econ.ucsd.edu/~rogordon/growth.218.doc>.
- [14] Podhradská, A. (2008): *Možnosti predikce vývoje daní*. Praha, Vysoká škola ekonomická, 2008. Bachelor thesis.
- [15] Ram, R. (1991): Elasticity of Individual Income Tax in the United States: Further Evidence from Cross-Section Data. [on-line], National Tax Journal, 1991, vol. 44, no. 1, pp. 93-99, [cited 8th August, 2013], <http://ntj.tax.org/wwtax%5Cntjrec.nsf/3302DAE258C2545A85256 86C00686D67/\$FILE/v44n1093.pdf>.
- [16] Špalek, J. Moravanský, D. (2005): Evaluace daňových predikcí v ČR. Finance a úvěr, 2005, vol. 55, no. 3-4, pp. 162-187.

Research of Estimates of Tax Revenue: An Overview

Ondřej BAYER

ABSTRACT

The paper deals with a summary of findings in the area of tax forecasts. It describes the basic methodology for predicting tax revenues, both in terms of the macroeconomic approach and from a microeconomic perspective. The microeconomic approach used microsimulation methods with methods based on tax elasticities. The macroeconomic approach explains the basic characteristics of regression analysis and time series analysis. It also describes methods for assessing the quality of predictions, together with the distribution of basic statistical error estimate.

Key words: Tax revenue; Forecasting; Microsimulations; Regression analysis.

JEL classification: H29.