

Mathematical Model of Income Tax Revenue on the UK example

Svetlana Ivanitskaya, Ilona V. Tregub

Financial University under the Government of Russian Federation
itregub@fa.ru

1. Description of the economic system. Formulation of the problem. Determination of endogenous and exogenous variables.

Econometrics is a field of economics that concerns itself with the application of mathematical statistics and is a tool of statistical inference to empirical measurement of relations postulated by economic theory. Thus, it helps analysing various economic phenomena, forecast their development and predict future fluctuations.

Creation of econometrical models provides an empirical evidence to support the evaluation of economic relationships, which occur in a modern society.

It is commonly known, that in every country all over the globe governments and central banks implement various policies in order to maintain the level of development, economic performance, monetary stability and standards of living. There are two main policies – fiscal and monetary. They are especially important when dealing with economic cycles consequences, like the one the whole world society is facing right now – after the financial crisis hit in 2008.

Fiscal policy is actions taken by the Government to influence the economy through the change in the revenues and expenditures of state budget. *The goals of fiscal policy* in short-run are to maintain the level of GDP; to achieve a stable price level and to get a full employment. *The instruments of the policy are:* Taxes (which influence both AD/AS shifts); Governmental spending (influence the shifts of AD curve) and Transfers (for both AD and AS changes).

There can be named different types of fiscal policy according to the state of the economy at a particular moment. Stimulating fiscal policy's aim is to reduce unemployment and to encourage economic activity at the period of a recession. To do so, government increases its spending (G), increases transfers as well (Tr) and reduces taxation (T_x). There is also a restraining fiscal policy, implemented in the period of economic expansion. The instruments are the same but each of them is used vice versa.

Fiscal policy can be either discrete or automatic. Discrete fiscal policy takes place when the government is changing state laws of tax rate, legal costs and so forth. Automatic policy is being implemented using the existing stabilisers: income tax, VAT, profit tax or unemployment benefits.

As far as every state's budget is concerned, taxes also play a huge role in it being one of the main components of the governmental revenue. Revenue side of fiscal mechanism must provide monetary base of governmental functions and also leave enough money for business and living needs.

Efficiency of revenue collection depends on careful complex structure and adjustments of various taxes, their bases, rates, exemptions and deductions. Optimal ratio of federal and local taxes creates adequate revenue sources for levels of authorities. General tax reform and increase of tax base or tax rate change invest-consume proportion in all sectors. Selective tax reform and tax exemption stimulate several preferred sectors and also damage tax fairness. That is why, analysing tax system is necessary. The topic is actual and important, that is why its current monitoring is significant.

The United Kingdom, being one of the leading countries in the modern world, is now facing obvious troubles dealing with its sovereign debt caused by the Global financial crisis. The British

government tries to come up with innovative ideas to reduce the amount of debt and federal budget deficit. And still taxes remain one of the instruments of the country's recovery policy implemented by the Parliament.

UK Income tax is forecast to raise £157.6 billion in 2012, but not all income is subject to tax. The primary forms of taxable income are earnings from employment, income from self-employment and unincorporated businesses, Jobseeker's Allowance, retirement pensions, income from property, bank and building society interest, and dividends on shares. Incomes from most means-tested social security benefits are not liable to income tax. Many non-means-tested benefits are taxable (e.g. the Basic State Pension), but some (notably Child Benefit) are not. Gifts to registered charities can be deducted from income for tax purposes, as can employer and employee pension contributions (up to an annual and a lifetime limit), although employee social security (National Insurance) contributions are not deducted. Income tax is also not paid on income from certain savings products, such as National Savings Certificates and Individual Savings Accounts.

This creative work is dedicated to provide an understanding of how taxes change depending on various factors. It represents an overview of the UK income taxation – the variables it reflects upon, recent fluctuations and forecast future changes occurrence.

Throughout this work, an econometrical model is created to forecast such variable as UK Income Tax revenue. This variable is **endogenous** – it depends on different factors.

Exogenous variables (independent factors, that build a base for relationship analysis) include:

- UK taxpaying population – the total number of people taxes are levied on;
- Inflation – as a reflection of the real value of money throughout different time periods;
- Oil prices – a variable changing the values of all economic indices the global economy analyses;
- GDP growth – representing the whole value of the economy, its current situation, economic condition and dynamics.

Econometric model built in this creative work lies on 3 main postulates:

1. Its specification is a result of a translation of economic phenomena provided in this section into math language.
2. Number of equations in the model is equal to the number of dependent variables, which is 1.
3. All variable used are dated – the statistics used are clearly specified in time periods.

The model is created using a general econometrics scheme.

1. Firstly, main economic laws, which are then analysed, are presented and explained.
2. Then the model is specified.
3. Afterwards all model parameters are estimated.
4. Then the entire model is tested.
5. If the model proves itself adequate, it then can be used to forecast economic phenomena described at the beginning.

All in all, it should help proving the existent relationship between endogenous and exogenous variables, stating economic laws sustainable and correct.

Every recession, and especially of such a scale as the one in 2008 proves to determine the relevant changes to improve and develop the financial sector of Great Britain. And that will lead to the improvement of the whole economy. Increasing GDP level already shows us that the UK is now in the stage of recovery, doing its best to provide further development for every branch of modern life, including politics, economy, finance and regulation. That is why estimating the relationships between British taxation, the country's population, the yearly level of inflation is of key importance in analysing the UK fiscal policy.

2. Description of statistical data related to the model

Endogenous variable

UK Income tax is the one which reflects tax revenue for the UK treasury and the government. The statistical data for this variable can be founded on the webpage of the HM Treasury, available at: <http://www.hmrc.gov.uk/thelibrary/national-statistics.htm> in the section of Tax Receipts and Tax payers, where there are xls and pdf documents concerning all taxes throughout available time period.

Exogenous variables (independent factors, that build a base for relationship analysis) include:

1. Quantity of taxpayers – the total number of people taxes are levied on;

Statistical data for this variable can be founded on the HM Treasury webpage as well, in the same section, regarding the same documents as for endogenous variable.

2. Inflation – as a reflection of the real value of money throughout different time periods.

One of the most reliable UK broadsheets The Guardian gives a perfect article representing all the relevant statistics from governmental sources. Retail price index is one of the components of the inflation analysis, taken to define current inflation in the country. The article is available at <http://www.guardian.co.uk/news/datablog/2009/mar/09/inflation-economics#data>

3. Oil prices – a variable changing the values of all economic indices the global economy analyses;

OECD webpage, where ilibrary is available, gives all the necessary statistics for the nominal and \$-adjusted oil prices for the chosen period in time:

<http://www.oecd-ilibrary.org/sites/factbook-2011-en/06/02/03/index.html;jsessionid=1gp3pg5q2xxjb.epsilon?contentType=/ns/Chapter./ns/StatisticalPublication&itemId=/content/chapter/factbook-2011-54-en&containerItemId=/content/serial/18147364&accessItemIds=&mimeType=text/html>

4. GDP growth – representing the whole value of the economy, its current situation and economic condition.

The perfect source for gathering a country's GDP rates is UNCTAD in its statistical branch UNCTADSTAT, available at:

<http://unctadstat.unctad.org/ReportFolders/reportFolders.aspx> Section Economical Trends, Subsection National Accounts.

In order to construct econometric model, firstly its necessary to provide a model specification – that is, to translate economic laws into mathematical language.

3.1. Model specification

As said in the previous section, UK income tax revenue depends on various factors – there are 4 of them.

Number of taxpayers, RPI, GDP growth rate and oil prices – these are regressors, which influence the endogenous variable Y – UK Income tax revenue.

Mathematically, the model should be as following:

Y – UK Income tax revenue

X₁ – Number of taxpayers

X₂ – UK annual RPI
 X₃ - world oil prices
 X₄ – UK GDP Growth

$$\left\{ \begin{array}{l} Y_t = a_0 + a_1X_{1t} + a_2X_{2t} + a_3X_{3t} + a_4X_{4t} + \varepsilon_t \\ \varepsilon(\varepsilon_t) = 0 \\ \sigma(\varepsilon_t) = \text{const} \end{array} \right.$$

where,

E is expectation of residual/disturbance term. Residual is, basically, a difference between real and theoretical points.

ε_t – is a disturbance term, showing random factors which can occur at any point in time and still influence the endogenous variable.

σ - is a standard deviation

Giving attention to all conditions of a model general form, the least square method should be applied.

Defining the dependence

In order to define the dependence of external factors, correlation matrix and scatter diagrams should be created.

Correlation matrix

To see, whether there is any correlation between all chosen variables, a correlation matrix should be constructed in Excel (*Image below*). Correlation matrix represents whether there occur a linear relationship between each exogenous variable and the explained variable. The bigger the correlation coefficient, the more linearly dependent is Y on a specific X. Correlation coefficient lies somewhere in between 0 and 1. So, if it is equal to 0, the variable are independent, while if it is equal to 1, the variable are perfectly dependent.

	Income tax revenue, mln GBP	# of taxpayers, 000s	GDP growth, %	RPI %	Oil Price, \$
Income tax revenue	1,00				
# of taxpayers, 000s	0,88	1,00			
GDP growth, %	0,08	-0,17	1,00		
RPI %	-0,59	-0,36	-0,33	1,00	
Oil Price, \$	0,47	0,73	-0,43	-0,04	1,00

A few words should be also said about the sign of each correlation coefficient in the given matrix. If the sign is positive, it indicates that there is a positive linear dependence. Conversely, if there is a “-” sign in front of the correlation coefficient, it represents a negative linear relationship between a particular causal variable and the Y.

Correlation matrix is a very useful tool, because not only can it tell the relationship between exogenous and endogenous variable, but it shows the dependence between each variable in response to all others presented in the matrix.

Scatter diagrams

It is useful to create scatter diagrams to graphically show, how the statistics for each variable are scattered through a trend line, representing the dependence of effect variable.

Here given 4 diagrams, representing 4 explanatory variables and Y response to the change of each of them.

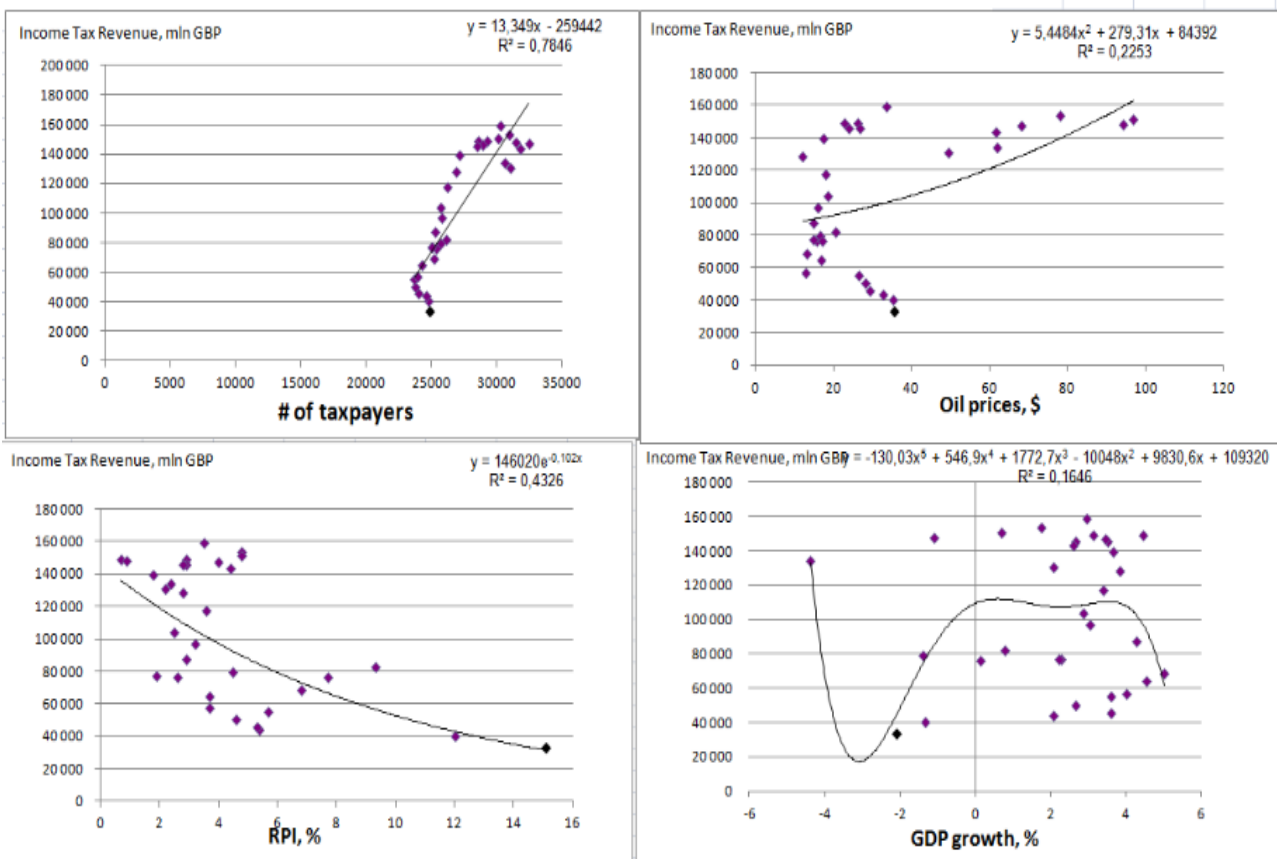
Diagram 1 represents the dependence of UK income tax revenue from the number of taxpayers.

Diagram 2 shows the relationship between UK income tax revenue and world oil price changes,

Diagram 3 shows how UK income tax revenue depends on Retail Price Index, a key inflation index.

Diagram 4 represents the absence of the relationship between UK income tax revenue and the rate of GDP growth. All point of this graph are scattered randomly, giving no possibility to define a distinct relationship between these two variables.

In all graphs a trend line is used to graphically show a precise dependence between each of exogenous variables and the dependent one. The equation of each trend line and determination coefficients also represent the strength and character of dependence between variables.



3.2. Estimation of coefficients in Excel

After proving the relationships between exogenous variables and the Y – UK Income Tax Revenue, it is now possible to estimate all coefficients in Excel by creating a regression econometric model.

It is done using Data Analysis facility. The input interval for x will be all the data collected for exogenous variable except the data for last year of 2011. The input data interval for Y will be all

the statistics for Income Tax revenue in the UK, again, without the data for the last year. We don't include last year's data, in order to eventually prove the model by trying to implement it on a given data.

The investigated econometric model can be written down as:

$$\left\{ \begin{array}{l}
 \mathbf{y}_t = -302970,53 + 11,71X_{1t} - 3,08X_{2t} - 3,86X_{3t} + 844,56X_{4t} + \mathbf{\epsilon}_t \\
 \quad (36156,92) \quad (1,41) \quad (886,66) \quad (184,87) \quad (1226,42) \quad (12367,49) \\
 R^2 \text{ adjusted} = 0,91 \\
 F = 77,95 \\
 F \text{ critical} = 2,74
 \end{array} \right.$$

After creation of a regression model in Excel, it becomes clear that some variables can be non-significant and thus, can be eliminated from the model. In a given example (See Appendix 1), several tests were done in order to define whether all variables X were important enough to be included in a model.

Firstly, R-tests proved the sufficient percentage of estimation of the dependent variable Y. In the given model $R^2=0,91$, which give a 91% probability that the forecast built on this model will be truth. That is, all the 4 Xs explain Y at 91%.

Then, in F-test, F critical was calculated by estimating the degree of freedom (regression df and residual df) and observation parameters. F critical should be compared to F calculated in a model by Excel. If the result of this comparison is $F \text{ calculated} > F \text{ critical}$, then determination coefficient R^2 is not random and thus, the model has a good quality of specification. In the given case:

$$F \text{ critical} = 2,74$$

$$F \text{ calculated} = 77,95$$

$F \text{ critical} < F \text{ calculated}$, thus the model proved itself trustworthy.

But T-test showed that, not all variables, which were included in the model in a very beginning, are significant. Here T critical was compared to T calculated. T critical is counted using observations and the probability of 5%.

$$T \text{ critical for a given model} = 2,06$$

Each of the t-statistics absolute values calculated for every variable X by Excel should be more than T critical in order to prove that corresponding regressors are important for a model.

t-statistics for $X_1 = 11,71$ (which is bigger than T critical – the variable is significant)

t-statistics for $X_2 = -3,08$ (the absolute value is higher than T critical – coefficient is important)

t-statistics for $X_3 = -3,86$ (the absolute value is higher than T critical – coefficient is important)

t-statistics for X_4 (GDP growth rate) in its absolute value was less than T critical, so the variable was proven non-significant and thus, has been eliminated from the model.

This way, the independent variable X_4 , which was the percentage rate of GDP growth, was eliminated from the model as non-significant.

A given example shows perfectly, how an independent variable, which can seem defining at first, can then prove itself insufficient for a given model, even though there might be some economical sense to put it in the model while starting a research.

A starting model, including all 4 variables X, can be found in Appendix 1. The final version of a model, which constitutes itself from one dependent and 3 independent variables, is given below.

31	2009	134093	30600	2,4	61,91	30664,31			
32	2010	150491	31000	4,8	78,08	31082,88			
33	2011	150939	30100	4,8	96,8	30201,6			
34									
35	ВЫВОД ИТОГОВ			Yth		127955,76			
36				Y+		153083,47			
37	Регрессионная статистика			Y-		102828,05			
38	Множественный R	0,96							
39	R-квадрат	0,92							
40	Нормированный R-к	0,91							
41	Стандартная ошибка	12246,48							
42	Наблюдения	31,00							
43									
44	Дисперсионный анализ			Fcritical		2,96			
45		<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Значимость F</i>			
46	Регрессия	3,00	47620357925,45	15873452641,82	105,84	0,00			
47	Остаток	27,00	4049362420,95	149976385,96					
48	Итого	30,00	51889720346,41						
49									
50		<i>Коэффициенты</i>	<i>Стандартная ошибка</i>	<i>t-статистика</i>	<i>P-Значение</i>	<i>Нижние 95%</i>	<i>Верхние 95%</i>	<i>Нижние 95,0%</i>	<i>Верхние 95,0%</i>
51	Y-пересечение	-302970,54	35714,44	-8,48	0,00	-376250,52	-229690,56	-376250,52	-229690,56
52	# of taxpayers, 000s	16,30	1,39	11,72	0,00	13,45	19,15	13,45	19,15
53	RPI %	2921,68	835,62	-3,50	0,00	-4636,22	-1207,13	-4636,22	-1207,13
54	Oil Price, \$	-781,37	169,66	-4,49	0,00	-1109,49	-413,25	-1109,49	-413,25
55			Tcritical	2,05					

3.3. Results interpretation

After creating a regression econometric model in Excel, it is now possible to interpret the figures given in all the table.

The Regression Statistics table represents:

R^2 – determination coefficient, showing the extent, to which variance of exogenous variables describes the variance of effect variable. The bigger the R^2 , the better is a model.

Standard error – represents a standard error of estimation, which is used during the stage of testing a model.

The Dispersion Analysis table shows:

- The coefficients used in a model for every variable included in econometrics model.
- Standard error of estimation of each coefficient – also used in writing a final equation in a model.
- F, T and P-statistics, used while testing the whole model and proving it sufficient with the help of various tests.
- The whole Dispersion Analysis table will be used in the following section, while estimating degrees of freedom, F critical, T critical and implementing all of model tests.

The final equation of an econometrical model can be given, using the results of the model construction:

$$y_t = -302970,54 + 16,3X_{1t} + 2921,68X_{2t} - 761,37X_{3t} + \epsilon_t$$

(35714,44)
(1,39)
(835,61)
(169,66)
(12246,48)

R^2 adjusted = 0,91
F = 105,84
F critical = 2,96

3.4. Tests: R²-Test, F-Test, T-Test, GQ-Test, DW-Test

In order to define, whether the model is trustworthy and applicable in real life, whether it can be forecasted upon and the least square method of estimation can be applied, the following tests were done.

R² shows the percentage of estimation of the dependent variable Y. In the given model R²=0,91, which give a 91% probability that the forecast built on this model will be truth. That is, all the 3 Xs explain Y at 91%.

In order to implement F-test, F critical should be calculated. It is done by estimating the degree of freedom (regression df and residual df) and observation parameters. Then, F critical should be compared to F calculated in a model by Excel. If the result of this comparison is F calculated > F critical, then determination coefficient R² is not random and thus, the model has a good quality of specification. In the given case:

$F_{critical} = 2,96$

$F_{calculated} = 105,84$

$F_{critical} < F_{calculated}$, thus the model is trustworthy.

T-test is used when it is necessary to find significant coefficients.

Here we again should compare T critical with T calculated. T critical is counted using observations and the probability of 5%.

T critical for a given model = 2,05

Each of the t-statistics absolute values calculated for every variable X by Excel should be more than T critical in order to prove that corresponding regressors are important for a model.

t-statistics for X1 = 11,72 (which is bigger than T critical – the variable is significant)

t-statistics for X2 = -3,50 (the absolute value is higher than T critical – coefficient is important)

t-statistics for X3 = -4,49 (the absolute value is higher than T critical – coefficient is important)

Goldfeld-Quandt Test

This test is done in order to check the residuals of a model and verify the condition of Gauss-Markov theorem. It helps checking homoskedasticity of the exogenous variables.

For implementing the GQ-test, all data must be sorted in ascending order of addition absolute values of Xs (Appendix 2: column Z).

Then the sorted data is divided into distinct data sets and they are analyzed respectively.

Regression models are built for separate data sets, in a similar way as for the initial model. It is clear, that F critical is calculated again for new data sets and this value is compared to GQ and 1/GQ values. It is important to understand, that $GQ = ESS1 / ESS2$ (ESS of a model for new data set 1) / ESS2 (ESS of a model for new data set 2).

If both values of GQ and 1/GQ are less than F critical, then it is proved that the residuals of an initial model are homoscedastic. And if so, then the least square method may be used in order to estimate econometrical model. Image shows, that

$$\begin{aligned} GQ &= 0,4 \\ 1/GQ &= 2,47 \\ F_{critical} &= 9,28 \end{aligned}$$

Thus, GQ-test is passed successfully – the model can be estimated through least square method.

Durbin-Watson Test

This test is dedicated to check the 3rd Gauss-Markov's theorem about the absence of autocorrelation between random residuals.

The algorithm of this test lies in the analysis of residuals, their lagged values. One should calculate the squared difference between the residual and its lag value for every observation, squared actual value of each residual and then find the relation between the sums of those. It is shown in the image below.

DW = 1.35 for this particular model.

Then we should check autocorrelation existence by distinguishing DW value into interval of critical values of Durbin-Watson statistics.

The observed value of Durbin-Watson (calculated on the basis of sample data) is compared with the critical value of Durbin-Watson, which is determined on a special table.

The critical value of Durbin-Watson is determined depending on the values of the upper and lower d_1 d_2 border criterion on special tables. These boundaries are defined, depending on the volume of sample n and the number of degrees of freedom $(h-1)$, where h - number of evaluated sample parameters (See image below).

If the observed value of the Durbin-Watson is less than the critical value of the lower boundary, $d < d_1$, then the fundamental hypothesis of no autocorrelation between the residuals of the regression model is rejected.

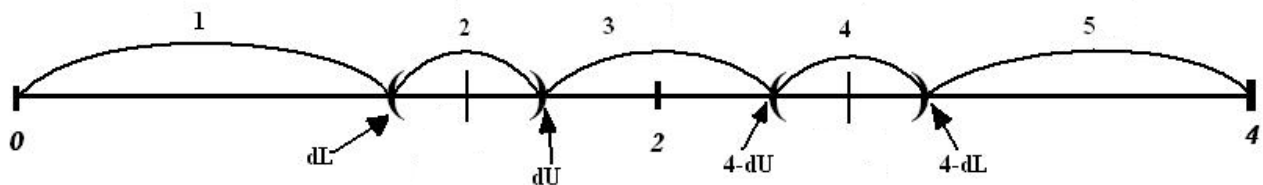
If the observed value of the Durbin-Watson more critical of its upper boundary, $d > d_2$, then the fundamental hypothesis of no first order autocorrelation between the residuals of the regression model is adopted.

If the observed value of the Durbin-Watson is between the upper and lower critical limits, that is, $d_1 < d < d_2$, adequate justification for making no single solution, further research is needed.

If the observed value of the Durbin-Watson more critical level $4 - d_1$, $d > 4 - d_1$, then the fundamental hypothesis of no first order autocorrelation between the residuals of the regression model is again rejected.

If the observed value of the Durbin-Watson is less than the critical value $4 - d_2$, $d < 4 - d_2$, then the fundamental hypothesis of no first order autocorrelation between the residuals of the regression model is once again adopted.

If the observed value of the Durbin-Watson is in the critical interval between the values of $4 - d_1$ and $4 - d_2$, is a sufficient basis for making the right decision is not only more research is needed.



$$\begin{aligned}
 DW &= 1,35 \\
 d_l &= 1,21 \\
 d_u &= 1,65 \\
 4-d_l &= 2,79 \\
 4-d_u &= 2,35
 \end{aligned}$$

In the investigated model, DW statistics lies in between d_l and d_u , which indicates that there is no information on existing autocorrelation between residuals. It makes a model applicable (See

Appendix 3). After implementation of this test, it is clear, that least square method can't be rejected.

3.5. Checking the adequacy of the model

Previous section explained, that the investigated model passes all necessary tests in order to prove its sustainability and suggest the implementation of the least square method. Now, the next step is to define, whether the model is applicable in forecasting. In order to do so, model's adequacy should be checked.

In order to check the adequacy of the model, we should take the last year's statistics and put them all into the model's equation, using all the coefficients from regression analysis. This would help calculating Y theoretical.

Y theoretical for the given model = 127955

After that, a confidence interval should be introduced in order to evaluate, whether real value of endogenous variable for the last years observed fits into that interval, built on the created model.

Confidence interval consists of higher and lower theoretical values of Y, each defined by the formula:

$$Y+ = Y_{\text{theoretical}} + T_{\text{critical}} * \text{Standard error} = 153083$$

$$Y- = Y_{\text{theoretical}} - T_{\text{critical}} * \text{Standard error} = 102828$$

It is clear that the real observed value of UK Income tax revenue for the year 2011 fits into the confidence interval. That means that the model is adequate and can be used for forecasting. (Appendix 4)

The last step before forecasting is estimating of a model error. It is calculated in a following way:

$$\sigma = |Y_{\text{theoretical}} - Y_{\text{real}}| / Y_{\text{real}} * 100\%$$

$$\sigma = |127955,756 - 150939| / 150939 * 100\% = 15\%$$

This means, that in 85% cases the model would give an exact right result.

3.6. Model forecasting

The created model is trustworthy and adequate, as proved in the previous two sections. It passed all the tests successfully, and represented itself as applicable for forecasting.

If use this model in forecasting, the future values of all exogenous variables should be extracted either by expert method or by estimating them as the effect variables. The future values of independent variables would help predict further fluctuations and dependencies of the endogenous variable Y – UK Income tax revenue.

4. Conclusions

Total UK government receipts are forecast to be £588.6 billion in 2012, or 38.1% of UK GDP. This is equivalent to roughly £12,000 for every adult in the UK, or £9,500 per person. Not all of this revenue comes from taxes: taxes as defined in the National Accounts are forecast to raise to £562.4 billion in 2012, with the remainder provided by surpluses of public sector industries, rent from state-owned properties and so on.

Income tax, National Insurance contributions and VAT are the largest sources of revenue for the UK government, together accounting for almost two-thirds of total tax revenue.

Econometrics model, created in this work, helps analyzing external factors, which the change in the UK Income tax revenue depends upon.

There are 3 factors influencing the revenue from UK Income tax.

First one is number of taxpayers. Here, the relationship between an exogenous and endogenous variable is clear - the more people pay taxes on their income, the more revenue UK government gets in form of tax receipts. So, if, for example, X_1 rises by 1000, income tax revenue will increase by 16 mln GBP per year.

Second independent variable is Retail Price Index, as a key inflation index. Inflation, obviously, influences the economy and volume of taxation. If inflation rises, money depreciates, thus increasing the average level of income taxation. So, if X_2 rose by 1%, income tax revenue for the UK government would increase by 2921 mln GBP in non-adjusted prices.

The third external factor influencing UK Income tax revenue, according to econometric model created in this work, is Oil Prices. Here, the dependence should be subjected to other different factors, like inflation. Because in some cases, the relationship between tax revenue and world oil prices can be negative. For example, low oil prices might be a reflection of high economic activity and, thus, increasing money supply. Then, income tax increase can be an element of contractional policy implemented in order to decrease money turnover and reduce the possibility of inflation. That's why in this case, the dependence tend to be negative.

All in all, it is necessary to state, that the creation of such an econometric model gives a great opportunity to try and forecast future fluctuations of endogenous variable (in the given case – UK Income tax revenue) to apply knowledge in fiscal and monetary policies adjustments, which is especially important in today's world of economic uncertainty.

Appendix

Appendix 1. A regression model including 4 independent variables.

Вывод итогов								
Регрессионная статистика								
Множественный R	0,96							
R-квадрат	0,92							
Нормированный	0,91							
Стандартная о	12367,49							
Наблюдения	31,00							
Дисперсионный анализ		F critical		2,74				
	df	SS	MS	F	Значимость F			
Регрессия	4,00	47692893092,85	11923223273,21	77,95	0,00			
Остаток	26,00	3976827253,56	152954894,37					
Итого	30,00	51669720346,41						
Кoefficients								
	Кoefficients	Стандартная ошибка	t-статистика	P-Значение	Нижние 95%	Верхние 95%	Нижние 95,0%	Верхние 95,0%
Y-пересечение	-304722,19	36156,92	-8,43	0,00	-379043,81	-230400,58	-379043,81	-230400,58
# of taxpayers, l	16,21	1,41	11,49	0,00	13,31	19,11	13,31	19,11
GDP growth, %	844,56	1226,42	0,69	0,50	-1676,38	3365,50	-1676,38	3365,50
RPI %	-2734,30	886,66	-3,08	0,00	-4556,85	-911,74	-4556,85	-911,74
Oil Price, \$	-713,56	184,87	-3,86	0,00	-1093,57	-333,55	-1093,57	-333,55
		T critical	2,06					

Appendix 2. GQ Test of a final regression model.

Year	Income tax	# of taxpayers, 0	RPI %	Oil Price, \$	Year	Income tax	# of taxpayers, 0	RPI %	Oil Price, \$		
1985,00	55438,00	23700,00	5,70	26,48	1990,00	82322,00	26100,00	9,30	20,50		
1984,00	50350,10	23800,00	4,60	28,24	1997,00	117633,00	26200,00	3,60	18,13		
1986,00	57116,00	23900,00	3,70	13,01	1998,00	128250,00	26900,00	2,80	12,16		
1983,00	45926,00	24000,00	5,30	29,39	1999,00	139384,00	27200,00	1,80	17,30		
1987,00	64509,00	24300,00	3,70	16,91	2003,00	145556,00	28500,00	2,80	26,76		
1982,00	43793,70	24600,00	5,40	32,82	2001,00	149112,00	28600,00	0,70	22,80		
1981,00	40318,30	24800,00	12,00	35,41	2002,00	145898,00	28900,00	2,90	23,85		
1980,00	32981,80	24900,00	15,10	35,85	2000,00	149084,78	29300,00	2,90	26,24		
1993,00	77271,00	25000,00	1,90	14,90	2011,00	150939,00	30100,00	4,80	96,80		
1989,00	76674,00	25000,00	7,70	15,68	2004,00	158975,00	30300,00	3,50	33,69		
1988,00	68813,00	25200,00	6,80	13,20	2009,00	134092,70	30600,00	2,40	61,91		
1994,00	87330,00	25300,00	2,90	14,76	2010,00	153491,00	31000,00	4,80	78,08		
1992,00	76346,00	25400,00	2,60	17,21	2005,00	130481,02	31100,00	2,20	49,54		
1991,00	79510,00	25700,00	4,50	16,56	2008,00	147855,58	31500,00	0,90	94,18		
1996,00	103892,00	25700,00	2,50	18,56	2006,00	143327,03	31800,00	4,40	61,52		
1995,00	97101,00	25800,00	3,20	16,09	2007,00	147323,51	32500,00	4,00	68,27		
Вывод итогов					Вывод итогов						
Регрессионная статистика					Регрессионная статистика						
Множесте	0,94				Множесте	0,78					
R-квадрат	0,88				R-квадрат	0,61					
Нормиро	0,85				Нормиро	0,51					
Стандарт	8165,09				Стандарт	12846,60					
Наблюд	16,00				Наблюд	16,00					
Дисперсионный анализ		F critical		3,49		Дисперсионный анализ		F critical		3,49	
	df	SS	MS	F	значимость F		df	SS	MS	F	значимость F
Регрессия	3,00	5675565783,27	1891855261,09	28,38	0,00	Регрессия	3,00	3113357581,63	1037785860,54	6,29	0,01
Остаток	12,00	800024926,18	66668743,85			Остаток	12,00	1980420457,62	165035038,13		
Итого	15,00	6475590709,45				Итого	15,00	5093778039,25			
Кoefficients					Кoefficients						
	Кoefficients	Стандартная ошибка	t-статистика	P-Значение	Нижние 95%	Верхние 95%	Нижние 95,0%	Верхние 95,0%			
Y-пересе	-296558,88	85283,06	-3,48	0,00	-482374,70	-110743,06	-482374,70	-110743,06			
# of taxpa	15,78	3,33	4,74	0,00	8,54	23,03	8,54	23,03			
RPI %	-2338,58	820,70	-2,85	0,01	-4126,73	-550,43	-4126,73	-550,43			
Oil Price, \$	-756,79	382,14	-1,98	0,07	-1589,39	75,82	-1589,39	75,82			
Y-пересе	13919,59	77631,12	0,18	0,86	-155224,09	183063,28	-155224,09	183063,28			
# of taxpa	4,78	2,79	1,72	0,11	-1,29	10,86	-1,29	10,86			
RPI %	-4492,48	1773,90	-2,53	0,03	-8357,49	-627,48	-8357,49	-627,48			
Oil Price, \$	-11,05	193,71	-0,06	0,96	-433,11	411,02	-433,11	411,02			

Appendix 3. DW test of a final regression model

93	DW	1,346788149
94	dl	1,21
95	du	1,65
96	4-dl	2,79
97	4-du	2,35
98		

Appendix 4. Checking adequacy of a final model. Confidence interval

The screenshot shows a Microsoft Excel spreadsheet with a regression analysis table. The table includes data for years 2000-2011, followed by summary statistics and regression coefficients.

	A	B	C	D	E	F	G	H	I	J
22	2000	149 085	29300	2,9	26,24	29329,14				
23	2001	149 112	28600	0,7	22,798	28623,50				
24	2002	145 898	28900	2,9	23,847	28926,75				
25	2003	145 556	28500	2,8	26,758	28529,56				
26	2004	158 975	30300	3,5	33,692	30337,19				
27	2005	130481	31100	2,2	49,536	31151,74				
28	2006	143327	31800	4,4	61,52	31865,92				
29	2007	147324	32500	4	68,27	32572,27				
30	2008	147856	31500	0,9	94,18	31595,08				
31	2009	134093	30600	2,4	61,91	30664,31				
32	2010	150491	31000	4,8	78,08	31082,88				
33	2011	150939	30100	4,8	96,8	30201,60				
34										
35	ВЫВОД ИТОГОВ			Yth	127955,76					
36				Y+	153083,47					
37	Регрессионная статистика			Y-	102828,05					
38	Множественный R	0,96								
39	R-квадрат	0,92								
40	Нормированный R-к	0,91								
41	Стандартная ошибк	12246,48								
42	Наблюдения	31,00								
43										
44	Дисперсионный анализ			Fcritical	2,96					
45		df	SS	MS	F	Значимость F				
46	Регрессия	3,00	47620357925,45	15873452641,82	105,84	0,00				
47	Остаток	27,00	4049362420,95	149976385,96						
48	Итого	30,00	51669720346,41							
49										
50		Козффициенты	Стандартная ошибка	t-статистика	P-Значение	Нижние 95%	Верхние 95%	Нижние 95,0%	Верхние 95,0%	
51	Y-пересечение	-302970,54	35714,44	-8,48	0,00	-376250,52	-229690,56	-376250,52	-229690,56	
52	# of taxpayers, 000s	16,30	1,39	11,72	0,00	13,45	19,15	13,45	19,15	
53	RPI %	2921,68	835,62	-3,50	0,00	-4636,22	-1207,13	-4636,22	-1207,13	
54	Oil Price, \$	-761,37	169,66	-4,49	0,00	-1109,49	-413,25	-1109,49	-413,25	
55			Тcritical	2,05						
56										
57										
58	ВЫВОД ОСТАТКА									