5/2023, sentyabr-oktyabr (№ 00066)

DEFINING THE VAT FRAUD COMPANY BY USING REGRESSION MODEL

Yusupov Jasurbek
Independent researcher of Fiscal Institute of Tax Committee, Uzbekistan. jasurbek_22@mail.ru

Ibragimov Boburshoh
Tashkent institute of finance, Uzbekistan Researcher of Fiscal Institute of Tax Committee, Uzbekistan. boburphd@gmail.com
DOI: https://doi.org/10.55439/EIT/vol11_iss5/a25

Abstract
Tax evasion is one of the factors that harm the economy, reducing budget revenues and creating an unequal competitive environment. One way to avoid paying value-added tax is to set up one-day companies in the shadow economy. The purpose of establishing these companies is to create an artificial value-added tax by companies that credit the value-added tax, reduce payments to the budget and, in the end, do not pay this tax to the budget. This article discusses the early detection of such intraday businesses using historical data. Based on the specific characteristics of one-day companies, when the company is newly established, it is possible to find an answer to the question of whether it will become a one-day company in the future or not through the probit regression model.
Keywords: VAT, tax evasion, one-day company

Introduction

The hidden economy is one of the measuring indicators of the informal sector of the economy, and it is a set of activities organized to avoid tax obligations or state control (Schneider, 2007). The high share of the hidden economy indicates the low efficiency of the tax system in these countries (Torgler & Alm, 2009). Tax avoidance and tax evasion a major focuses of tax administration, especially in developing countries (Davia, Coggins, Wideman, & Kastantin, 2000).

Value-added tax is introduced in 170 countries of the world, and the share of this tax is a significant contribution to the state budget. In Uzbekistan, the share of VAT in the state budget has exceeded 30 percent since 2018, and its share in the GDP exceeded 7 percent by 2021 (Table 1).

The increase in value-added tax revenue is also related to the increase in the number of taxpayers. The increase in the number of taxpayers reduces the ability of the tax administration to control the activities of taxpayers and creates the opportunity for taxpayers to use various tax avoidance schemes. One of these opportunities is to form one-day companies in the hidden economy, to reduce the amount of VAT due to the budget by

creating fake invoices. This can be more easily understood in the following conditional example scheme:

Table 1

Macroeconomic indicators and VAT revenues of the Republic of Uzbekistan for 2017-2021.

Indicators	2017	2018	2019	2020	2021
GDP	317,5	426,6	532,7	605,5	738,4
Budget revenues	49,7	79,1	112,2	132,9	164,7
VAT	14,7	27,9	37,1	40,2	52,8
Share of VAT in GDP	4,6	6,5	7,0	6,6	7,2
Share of VAT in total revenue	29,6	35,2	33,1	30,2	32,1

Note: Data from the Statistics Agency and the Ministry of Economy and Finance

In this scheme, the operations up to the sale of apple juice to the final customer as a finished product are shown. In the conditional example, since an apple-growing household was considered, VAT was not calculated in the budget. He sold his product to a wholesale company for 100 conditional units. The seller sold the product to the manufacturer for 120 conditional units, putting a 20 percent margin on the received product. At the same time, VAT on the payment of 18 conditional units appeared in the seller company at the rate of 15 percent of VAT. The manufacturer produced apple juice from the obtained apple product and sold its product to the final customer for 150 conditional units. From this amount, 22.5 conditional unit was calculated for VAT at the rate of 15 percent. The manufacturer has the right to credit the 18 conditional units paid to the seller when receiving apple products. Therefore, the manufacturer pays 4.5 (22.5-18) conditional units to the budget. In the end, a total of 22.5 conditional units should go to the budget in this chain.

Scheme of juice production and sale to final consumers

But in this case, what if the supplier or vendor was established as a one-day company? That is, the goal is to avoid paying taxes, issue invoices, and ultimately not pay to the budget. Since the manufacturer has submitted an invoice for the 18 conditional units, the manufacturer has right to credit this amount. Although this amount has not been paid to the budget by the selling company, there is no legal limit to this. In this case, instead of the 22.5 conditional units that should go to the budget, 4.5 conditional units will be received.

The above example is just one way to avoid paying VAT, and since such a situation is common in Uzbekistan, the tax administration conducts an uncompromising fight against one-day companies.

In this study, based on the previous operations and specific characteristics of such companies, the probability of becoming a one-day company in the future is calculated when the company is newly established. This in turn makes it easier for the tax administration to perform manual tasks such as analyzing the transactions of the company through bank statements or invoices after the completion of its activities.

Literature review

In the research conducted by Gonzalez and Velázquez (2012), using various deep data analysis methods, based on the tax payments, previous operations and characteristics of the companies, they estimated the probability of these companies issuing false invoices. In particular, self-organizing maps, neural gas, classification trees, and multilayer perceptron neural network methods of deep data analysis were used by them.

Mehta et al. (2019) used machine learning and statistical analysis of taxpayers in Telangana, India to identify suspicious businesses in their study.

Zao and Bai (2022) investigated the possibility of financial suspicious transactions by listed companies using machine learning algorithms. Specifically, they modelled 18,060 transactions based on the 13 most influential indicators.

Gupta and Nagadevara (2007) developed a list of tax evasion companies through in-depth data analysis to identify them for a tax audit. This allowed the tax administration to effectively use resources in identifying companies that should be audited.

Model and methods

We use the probit regression model to determine the one-day company, and the model looks like this:

$$P(ODC = 1|X) = \Phi(b_0 + b_n X)$$

Here, **ODC** is the dependent variable and represents one-day company and is a binary number. If the company is suspicious then it equals 1, if it is a reliable company, then it equals 0.

X is a vector of independent variables consisting of:

$\mathbf{X}_1$ - the director of the company worked as a director in several other companies. In the historical data on companies engaged in the secret economy, it was found that the heads of many one-day companies were previously directors of shadow companies. Therefore, according to our hypothesis, this independent variable has a positive correlation with the dependent variable.

$\mathbf{X}_2$ - how many companies did the director of the company open in one day. In practice, it was found that the directors of some one-day companies opened up to 15 companies in one day. Therefore, according to our model, this independent variable will have a positive correlation with the dependent variable according to our hypothesis.

$\mathbf{X}_3$ - the number of real estate owned by the company. Usually, when a business is established for the purpose of a one-day business, these businesses do not have any tangible or nontangible assets. Because these companies will end up incurring a large amount of debt to the budget and will cease their activities, the expropriation will be focused on their existing assets. According to our hypothesis, this variable has a negative correlation with the dependent variable.

$\mathbf{X}_4$ - the number of real estate owned by the founder of the company. As with our variable above, the presence of property of the founder, not the director of the company, increases the probability that the debt arising at the end of the activity will be collected from

the founder's property. Therefore, according to our hypothesis, this variable also has a negative correlation with the dependent variable.

X_5 - the number of employees in the company. Historically, one-day businesses are typically run by one or two people. According to our hypothesis, the increase in the number of employees has a negative correlation with the dependent variable.

X_6 - the number of "zero" reports submitted by the company to the personal income tax. The purpose of including this indicator in the model is that usually one-day companies are newly established and report "zero" until they start their operations, then they make a large turnover in a short period of time and then they are observed to "disappear". Therefore, according to our hypothesis, the high of this indicator has a positive correlation with the dependent variable.

X_7 - the active period of the company. As per the assumption in our previous variable, one-day businesses are newly established businesses. The difference between the date of last operation of the company and the date of state registration indicates how long it has been in existence. According to our hypothesis, this indicator has a negative correlation with the dependent variable.

X_8 - the share of canceled online electronic invoices in total issued invoices. Analysis shows that there are many cases of one-day companies issuing invoices first and then canceling them. For this reason, this indicator is also included in the model, and the dependent variable is assumed to have a positive correlation.

X_9 – the turnover amount (including VAT) corresponding to one invoice. Historically, one-day businesses have recorded large turnover in a short period of time, reflecting this large turnover on a small number of invoices. Therefore, a large amount issued in one invoice has a positive correlation with the dependent variable according to our hypothesis.

X_{10} – the existence of online cash register (1-exists, 0-not exists). Typically, one-day businesses are more common in the wholesale and construction industries and less common in the retail sphere. Therefore, the fact that the company is related to the retail sector and the presence of an online cash register reduces the possibility of its future transition to the hidden activities.

X_{11} - the difference (in days) between the created and formed dates of online electronic invoice. The proximity of this indicator to "zero" means that the company issues invoices on time, on the date of realization. A higher rate indicates that the company is issuing invoices later than on time. According to our hypothesis, this indicator also has a positive correlation with the dependent variable.

In general, the sign of all independent variables in our model according to our hypothesis is presented in the table below (*Table 2*).

Data selection and descriptive statistics

Currently, in the tax system, a daily analysis is carried out on the basis of the bank transfers and formed invoices of the legal entities in the determination of one-day companies. If suspicious transactions, i.e. sale of goods that have not been received, sale of highly liquid goods in large quantities, and similar cases are studied, companies carrying out suspicious transactions are isolated. After that, the activities of these companies are analyzed in depth, a one-day company list is formed and published on the website of the tax authority.

Table 2

An initial assumption about the effect of the independent variable on the dependent variable.

№	Independent variable	An effect on dependent variable (Assumption)	An effect on dependent variable (Fact)	Comparing the results
1	The director of the company worked as a director in several other companies	Positive	Positive	✓
2	How many companies did the director of the company open in one day	Positive	Positive	✓
3	The number of real estates owned by the company	Negative	Negative	✓
4	The number of real estates owned by the founder of the company	Negative	Negative	✓
5	The number of employees in the company	Negative	Negative	✓
6	The number of "zero" reports submitted by the company to the personal income tax	Positive	Positive	✓
7	The active period of the company	Negative	Negative	✓
8	The share of canceled online electronic invoices in total issued invoices (%)	Positive	Positive	✓
9	The turnover amount (including VAT) corresponding to one invoice	Positive	Positive	✓
10	The existence of online cash register (1-exists, 0-not exists)	Negative	Negative	✓
11	The difference (in days) between the created and formed dates of online electronic invoice	Positive	Positive	✓

During the selection of one-day companies, 500 companies from 2690 companies in this list and 500 companies from other companies not included in the list were randomly selected as reliable companies. Some of the selected companies were excluded from the sample due to the fact that they were newly established, did not fully operate, and had incomplete turnover. Then, in the end, a total of 900 companies remained in the sample, of which 499 were suspicious companies and 401 were reliable companies (*Table 3*).

Table 3

Descriptive statistics about reliable and suspicious businesses.

Panel A: Summary statistics for total samples					
Variables	*Sample*	*Mean*	*Standard deviation*	*Minimum*	*Maximum*
ODC	900	0.5544	0.4973	0	1
X_1	900	3.6778	3.3399	0	34
X_2	900	1.2756	1.0795	1	15
X_3	900	0.4689	1.2209	0	24
X_4	900	0.4900	1.2878	0	24
X_5	900	18,1956	177.4579	0	4393
X_6	900	1.8600	3.3074	0	19
X_7	900	300,1600	402.6666	0	5713
X_8	900	14.0322	15.6985	0	100
X_9	900	151.0367	667,1085	0	15335
X_{10}	900	0.2189	0.4137	0	1
X_{11}	900	32.7011	177.3203	0	4135

Panel B: Statistics on suspicious companies					
Variables	*Sample*	*Mean*	*Standard deviation*	*Minimum*	*Maximum*
ODC	499	1	0	1	1
X_1	499	4.6413	4.1393	0	34
X_2	499	1.4770	1.4046	1	15
X_3	499	0.1784	0.5435	0	5
X_4	499	0.1663	0.5510	0	5
X_5	499	1.6994	3.6523	0	40
X_6	499	2.7976	3.9618	0	19
X_7	499	71.7595	94.6249	0	730
X_8	499	14.3046	17.8504	0	100
X_9	499	195.7756	874.2549	0	15335
X_{10}	499	0.0240	0.1534	0	1
X_{11}	499	6.3026	19.6188	0	229
Panel V: Statistics on reliable companies					
Variables	*Sample*	*Mean*	*Standard deviation*	*Minimum*	*Maximum*
ODC	401	0	0	0	0
X_1	401	2.4788	1.0678	0	5
X_2	401	1.0249	0.2222	1	3
X_3	401	0.8304	1.6572	0	24
X_4	401	0.8928	1.7481	0	24
X_5	401	38.7232	264.5718	3	4393
X_6	401	0.6933	1.6072	0	9
X_7	401	584.3791	455,1950	0	5713
X_8	401	13.6933	12.5287	0	48
X_9	401	95.3641	207.4340	0	1840
X_{10}	401	0.4613	0.4991	0	1
X_{11}	401	65.5511	261,2213	0	4135

Table 4

Correlation table of variables

Indicators	**ODC**	**X_1**	**X_2**	**X_3**	**X_4**	**X_5**	**X_6**	**X_7**	**X_8**	**X_9**	**X_{10}**	**X_{11}**
ODC	1.0000											
X_1	0.3220	1.0000										
X_2	0.2082	0.5198	1.0000									
X_3	-0.2656	-0.0537	-0.0745	1.0000								
X_4	-0.2805	-0.0832	-0.0860	0.8668	1.0000							
X_5	-0.1038	-0.0292	-0.0234	0.0600	0.0035	1.0000						
X_6	0.3164	0.1801	0.0510	-0.0788	-0.1066	-0.0450	1.0000					
X_7	-0.6331	-0.2049	-0.1386	0.2003	0.2026	0.0710	-0.2003	1.0000				
X_8	0.0194	0.0058	-0.0589	0.0083	0.0064	0.0204	-0.0054	0.0575	1.0000			
X_9	0.0749	-0.0032	-0.0132	-0.0115	-0.0063	0.0010	-0.0213	-0.0556	0.0275	1.0000		
X_{10}	-0.5256	-0.1832	-0.0904	0.2084	0.2536	-0.0192	-0.1743	0.3514	-0.0365	-0.0816	1.0000	
X_{11}	-0.1662	-0.0546	-0.0400	0.0615	0.0699	0.0883	-0.0750	0.1544	-0.0307	-0.0250	0.0745	1.0000

In this correlation table, almost all the independent variables have the same relationship with the dependent variable as we assumed. Only the difference between the date the electronic invoice was created and issued shows a negative correlation, not a positive one, as we hypothesized. This correlation is also compared with the results after running the model in the next section.

Results

According to the above model, we first regressed our dependent variable separately with each independent variable using a probit model. Then, a regression was performed including all independent variables in the model and the results in the table below were obtained (*Table 5*).

Table 5

Results of regression of the dependent variable with each of the independent variables separately and jointly.

Independent variables	*ODC*	*ODC*	*ODC*	*ODC*	*ODC*	*ODC*	*ODC*	*ODC*	*ODC*	*ODC*	*ODC*	*ODC*
	1	2	3	4	5	6	7	8	9	10	11	All
X1	0.2369***											0.1880***
	(9.25)											(2.80)
X2		0.6082***										0.0338
		(5.36)										(0.12)
X3			-0.5621***									-0.3064*
			(-8.99)									(-1.70)
X4				-0.5678***								-0.3140*
				(-9.17)								(-1.78)
X5					-0.1594***							-0.0859***
					(-14.69)							(-5.78)
X6						0.1779***						0.1373***
						(8.86)						(3.33)
X7							-0.0067***					-0.0069***
							(-20.59)					(-12.21)
X8								0.0016				0.0385***
								(0.59)				(4.34)
X9									0.0003**			0.0009*
									(2.35)			(1.88)
X10										-2.0508***		-0.7943***
										(-13.69)		(-2.75)
X11											-0.0157***	0.0013**
											(-8.29)	(2.54)
Intercept	-0.6203***	-0.5605***	0.3637***	0.3693***	0.8385***	-0.1289***	1.9855***	0.1147**	0.0946**	0.5036***	0.3804***	1.3977***
	(-7.26)	(-4.37)	(7.49)	(7.63)	(13.45)	(-2.59)	(19.00)	(2.03)	(2.10)	(10.17)	(7.75)	(3.92)
Number of companies	900	900	900	900	900	900	900	900	900	900	900	900

Robust z-statistics in parantheses

**** p < 0.01, ** p < 0.05, * p < 0.1*

When we perform a separate regression with each independent variable, the coefficients of almost all variables except for variable 8 are statistically significant within the 1% range. X_8 - the interaction of the independent variable with the dependent variable in a separate case is not statistically significant, but when running the model with all independent variables, it is statistically significant in the range of up to 1%.

In addition, when X_2 - the indicator of how many companies the company director opened in one day is regressed separately with the dependent variable, it is statistically significant in the range of up to 1%, but when using the model with all independent variables, only this indicator loses statistical significance.

In general, since only one of the 11 independent variables had a coefficient that was not statistically significant, the indicators were re-predicted based on the results of the model and the results were compared with the actual situation in the sample. The results and comparison table are presented (*Table 6*).

Table 6

Prediction of sample rates using the model and comparison with actual rates

In the sample company status	Model company status	Companies according to the model the number	For example relative model indicator share *(in percent)*	Explanation
Reliable		401	100	
Reliable	Reliable	379	95	True
Reliable	Suspicious	22	5	False
Suspicious		499	100	
Suspicious	Suspicious	483	97	True
Suspicious	Reliable	16	3	False

Source: correctly rated overnight businesses as suspicious 97 percent of the time.

Discussion and conclusions

If we compare the results of the model with our assumptions at the beginning of the study, we will have the following comparison table:

Table 7

Comparison of initial hypothesis and model parameters

№	Independent variables	The effect on dependent variable (Assumption)	The effect on dependent variable (Fact)	Comparing the results
1	The director of the company worked as a director in several other companies	Positive	Positive	✓
2	How many companies did the director of the company open in one day	Positive	Positive	✓
3	The number of real estates owned by the company	Negative	Negative	✓
4	The number of real estates owned by the founder of the company	Negative	Negative	✓
5	The number of employees in the company	Negative	Negative	✓
6	The number of "zero" reports submitted by the company to the personal income tax	Positive	Positive	✓
7	The active period of the company	Negative	Negative	✓
8	The share of canceled online electronic invoices in total issued invoices (%)	Positive	Positive	✓
9	The turnover amount (including VAT) corresponding to one invoice	Positive	Positive	✓
10	The existence of online cash register (1-exists, 0-not exists)	Negative	Negative	✓
11	The difference (in days) between the created and formed dates of online electronic invoice	Positive	Positive	✓

The result shows that although the coefficient of our 2nd independent variable was not statistically significant, the correlation of all variables with the dependent variable gave the same result as our hypothesis.

At the next stage, a simulation was carried out using the model and including some parameters (*Appendix 1*). As a result of the simulation, when intra-day firm-specific and non-specific indicators were included in the model, the model also produced the results we expected.

In general, this study is the first step in determining the day-to-day companies engaged in secret activities in Uzbekistan using the regression model, and requires more in-depth study. In particular, it includes changes such as increasing the number of samples, revising the independent variables based on the characteristics of one-day companies, using the identification codes of the implemented products and services, and including the indicators for determining the imbalance of goods in the model. The development of this area allows the tax administration to use resources effectively, reducing the human factor and manual labor.

List of used literature

1. Alm , J., & Torgler , B. (2011). Do ethics matter? Tax compliance and morality. *Journal of Business Ethics* , *101* , 635-651.

2. Assylbekov , Z., Melnykov , I., Bekishev , R., Baltabayeva , A., Bissengaliyeva , D., & Mamlin , E. (2016). Detecting value-added tax evasion by business entities of Kazakhstan. In *Intelligent Decision Technologies 2016: Proceedings of the 8th KES International Conference on Intelligent Decision Technologies (KES-IDT 2016)–Part I* (pp. 37-49). Springer International Publishing .

3. Davia , H., Coggins, P., Wideman, J., & Kastanin , J. (2006). Accountant's guide to fraud detection and control.

4. González, PC, & Velásquez , JD (2013). Characterization and detection of taxpayers with false invoices using data mining techniques. *Expert Systems with Applications* , *40* (5), 1427-1436.

5. Gupta, M., & Nagadevara , V. (2007, December). Audit selection strategy for improving tax compliance: application of data mining techniques. In *Foundations of Risk-Based Audits. Proceedings of the eleventh International Conference on e-Governance, Hyderabad, India, December* (pp. 28-30).\

6. Krumplytė , J. (2010). The shadow economy factors and causes analysis by expert evaluation method: case of Lithuania. *Business , Management oath Economics Engineering* , *8* (1), 122-138.

7. Mehta, P., Mathews, J., Kumar, S., Suryamukhi , K., Babu , CS, Rao, SKV, ... & Bisht, D. (2019). Big data analytics for tax administration. In *Electronic Government and the Information Systems Perspective: 8th International Conference, EGOVIS 2019, Linz, Austria, August 26–29, 2019, Proceedings 8* (pp. 47-57). Springer International Publishing .

8. Buehn , A., & Schneider, F. (2007). Shadow economies and corruption all over the world: Revised estimates for 120 countries. *Economics : The Open-Access , Open-Assessment E- Journal* , *1* (9), 1-66.

9. Stankevicius , E., & Leonas , L. (2015). Hybrid approach model for prevention of tax evasion and fraud. *Procedia-Social oath Behavioral Sciences* , *213* , 383-389.

10. Zhao, Z., & Bai, T. (2022). Financial Fraud Detection and Prediction in Listed Companies Using SMOTE and Machine Learning Algorithms. *Entropy* , *24* (8), 1157.

Appendix

One-day company identification simulation using the probit model

Example 1: Indicators

X $_1$ - How many companies does the head of the company lead - **5**

X $_2$ - How many companies did the head of the company open in one day - **3**

X $_3$ - The number of real estates in the name of the head of the company - **0**

X $_4$ - The number of real estates in the name of the founder of the company - **0**

X $_5$ - The number of employees in the company - **1**

X $_6$ - Number of "zero" reports submitted to income tax - **0**

X $_7$ - The difference between the date of the last operation and the date of registration - **20 days**

X $_8$ - Share of canceled EHFs in total issued EHFs – **2%**

X $_9$ - the amount of turnover corresponding to one invoice for EHFs **- 150 million soums**

X $_{10}$ – Presence of online NKM – **0**

X $_{11}$ – the difference between the date of EHF and the date of creation of EHF – **30 days**

The indicator according to the model is **0.99321967 ≈ 1**

Example 2: Indicators

X $_1$ - How many companies does the head of the company lead - **2**

X $_2$ - How many companies did the head of the company open in one day - **0**

X $_3$ - The number of real estate in the name of the head of the company - **1**

X $_4$ - Number of real estate in the name of the founder of the company - **2**

X $_5$ - The number of employees in the company - **10**

X $_6$ - Number of "zero" reports submitted to income tax - **0**

X $_7$ - The difference between the date of the last operation and the date of registration - **600 days**

X $_8$ - Share of cancelled EHFs in total issued EHFs – **1%**

X $_9$ - the amount of turnover corresponding to one invoice for EHFs **- 40 million soums**

X $_{10}$ – Presence of online NKM – **1**

X $_{11}$ – Difference between EHF date and EHF creation date – **0 days**

The indicator according to the model is 0.000004799 ≈ 0