Research on Indicators of Tax Assessment Based on VAT Invoice

Zhong-ping LIU¹ and Xiao-yuan LIU²

¹College of Computer Science and Engineering, Yulin Normal University, Yulin, Guangxi, China
²Library, Yulin Normal University, Yulin, Guangxi, China

Keywords: Invoice, Tax assessment, Assessment indicator, Case-selected, Indicator threshold.

Abstract. The method of deducting tax in value-added tax (VAT) is to indicate the tax in the invoice. VAT invoice is not only the important proof of taxpayer's economic activity, but also the legal proof of the seller's tax obligation and the buyer's input tax deduct. Tax assessment refers to a management activity of tax authorities in order to identify the authenticity and legality of taxpayers' declaration. The purpose of this paper is to provide case-selected of tax assessment through the analysis of VAT invoices. Firstly, this paper analyzed the present situation of tax assessment, and pointed out that the man-machine combination method should be the main method of tax assessment in China. Secondly, five kinds assessment indicators were put forward based on VAT invoice. Finally, through the calculation of a large number of invoice data, the thresholds of all indicators were determined. Each indicator threshold provided the basis for the case-selected of tax assessment.

Introduction

Value-added tax (VAT) is a turnover tax based on the increment of goods (including taxable services) in the course of circulation. VAT has become one of the most important taxes in China. The special invoice of VAT is not only the important accounting evidence of the taxpayer reflecting economic activities, but also the legal proof of the tax obligation of the seller and the tax amount of the buyer[1]. It is also an important and decisive legal special invoice in the calculation and management of VAT. Tax assessment is one of the ways of tax administration and tax service. Tax assessment plays an important role in helping taxpayers to discover and correct the mistakes and omissions in the process of fulfilling their tax obligations, and to correct the taxpayers' consciousness of paying taxes.

Summary of Research Status

In 2005, on the basis of conceptual disputes, the General Administration of Taxation gave a new definition of tax assessment in the Administration of tax Assessment (trial): Tax assessment means that the tax authorities use the method of comparative analysis of data to make qualitative and quantitative judgments on the authenticity and accuracy of tax returns of taxpayers or withhold agents and to take further management measures. In terms of location, tax assessment is between daily tax declaration and tax inspection, and its role is similar to that of filter screen. In 2014, the State Administration of Taxation issued "Opinions on Strengthening Tax Risk Management" and "Tax Service and Management Rules for Large Enterprises (trial)", which identified tax assessment as an important means to deal with risks in tax risk management [2].

In some developed countries and regions, due to the high level of information, they have established a complete assessment system relying on computers. In recent years, the informative level of our country has also been greatly improved. With the development of the gold tax project, our tax authorities have also started to actively study the advanced tax assessment methods and means from abroad[3]. At present, the common means of tax assessment in China are check method and analysis method. The check method is to judge the reasonableness of declare data by analyzing and checking or other means. The analysis method is to study the Linkages among indicators and aggregate analysis the data from both internal and external sources.
Tax assessment can be divided into manual method, computer method and human-machine combination method[4]. Manual methods are usually used to judge unconventional doubts or risks. Computer method has the advantage of high efficiency. The method can generate indicators for the corresponding objects through the data in the Database and can assessment declarations taking it as a reference. The human-machine combination method refers to generate selected-cases by computer system and then analysis by manual. The aim of this paper is to mark possible abnormal taxpayers through dynamic monitoring of VAT invoices and to provide selected-cases for further tax assessment.

Structure of Assessment Indicators

VAT is a kind of turnover tax levied on the basis of the sales of goods and taxable services and applying the principle of tax deduction [5]. Since the certification period of VAT is 3 months, the calculation of our indexes is not based on months, but on sliding windows with a span of 3 months. From the value added tax invoice, we can build the following indicators.

1) VAT tax burden rate
   - Meaning: Tax burden rate refers to the ratio of the value added tax payable to the taxable income. The tax burden rate of VAT is the most common and widely used index in tax assessment.

   $$\text{TBR}=\frac{\text{CTP}}{\text{CTI}}\times 100\%$$
   (TBR: Tax burden rate; CTP: Current tax payable; CTI: Current taxable income)

   - Analysis: This indicator can be compared with the warning value.
   - Realization: For the convenience of calculation, we use the deviation from the mean of the industry and the fluctuation compared historical data to characterize the change in the tax burden.

   $$\text{TBR}_\text{Deviation}=\frac{\text{TBR}-\text{MTBR}}{\text{MTBR}}$$
   (MTBR: mean of the industry TBR)

   $$\text{TBR}_\text{Fluctuation}=\frac{\text{CTBR}-\text{PTBR}}{\text{PTBR}}$$
   (CTBR: Current tax burden rate; PTBR: Previous tax burden rate)

2) Sales growth rate
   - Meaning: The amount of sales tax depends on the amount of sales income. The growth rate of sales tax reflects the change of sales income, which can be used to judge whether the tax amount of sales is abnormal.

   $$\text{SGR}=\frac{\text{CSI}-\text{PSI}}{\text{PSI}}$$
   (SGR: Sales growth rate; CSI: Current sales income; PSI: Previous sales income)

   - Analysis: Under normal circumstances, the growth rate of income is basically the same as the growth rate of industry income and the growth rate of self tax.
   - Realization: we use the following two indicators relating to sales growth rates to measure whether an enterprise is abnormal.

   $$\text{SGR}_\text{Deviation}=\frac{\text{SGR}-\text{MSGR}}{\text{MSGR}}$$
   (MSGR: mean of the industry SGR)

   $$\frac{T&I}{\text{Consistency}}=\frac{\text{TGR}}{\text{IGR}}$$
   (TGR: Tax growth rate; IGR: incomes growth rate)

3) Invoice quantity growth rate
   - Meaning: Invoice quantity is the sum of normal invoice, red letter invoice and scrap invoice. If the invoice quantity increases suddenly and substantially, then the enterprise may have anomalies.

   $$\text{IQG}=\frac{\text{CIQ}-\text{PIQ}}{\text{PIQ}}\times 100\%$$
(IQG: Invoice quantity growth rate; CIQ: Current invoice quantity; PIQ: Previous invoice quantity)

- **Analysis:** If the amount of invoice exceeds the limit, you should apply to the tax department for purchase and be subject to review. Therefore, we do not need to forecast this situation. Our goal is to find those that do not exceed the limit, but which fluctuate abnormally.

- **Realization:** It is normal to have large fluctuations in the case of low invoice usage, so it is not appropriate to use the growth rate alone. So we designed this indicator as an early warning marker:

\[
\text{IQ\_Label} = \begin{cases} 
1 & (IQG \geq 30\%) \text{and} (IQ \geq 10) \\
0 & (IQG < 30\%) \text{ or } (IQ < 10)
\end{cases}
\]

\[(8)\]

**4) Invalid invoice quantity**

- **Meaning:** There are four common cases of invoice invalidation: invoicing errors, return of sales, discount and termination of service. In the database, invalid invoices are marked with invalid signs.

\[\text{IIQ} = \text{CIIQ} \] \hspace{1cm} \text{(9)}

\[(IIQ: \text{Invalid invoice quantity}; \text{CIIQ: Current invalid invoice quantity)}\]

- **Analysis:** The excessive number of invalid invoice may be due to the following abnormal reasons: malicious concealment of income, malicious collusion or virtual invoicing.

- **Realization:** We use the invalid invoice rate instead of the invalid invoice quantity.

\[\text{II\_Rate} = \frac{\text{CII}}{\text{CIQ}} \] \hspace{1cm} \text{(10)}

\[(II\_Rate: \text{Invalid invoice rate}; \text{CII: Current invalid invoice}; \text{CIQ: Current invalid quantity})\]

**5) Red letter invoice quantity**

- **Meaning:** Non-current invoice to be canceled (more than one month) can only be hedged with red letter invoice. In the database, the red letter invoice is represented as a negative number invoice.

\[\text{RLIQ} = \text{CRLIQ} \] \hspace{1cm} \text{(11)}

\[(RLIQ: \text{Red letter invoice quantity}; \text{CRLIQ: Current red letter invoice quantity})\]

- **Realization:** Since the number of red letter invoices is very small, we use Eq.11 as an indicator.

**Determination of Indicator Thresholds**

This paper takes the VAT invoice of an industry in a province as a sample to study the determination of each indicator threshold.

All the data is placed in three tables. The taxpayer information table include industry code, taxpayer ID, registration type code, registration date, modification date and abnormal label. The invoice content table include invoice ID, seller ID, buyer ID, sales income, tax, total of income and tax, invoice month, invoice date, cancel label. The invoice detail table include invoice ID, invoice month, description of goods, specifications and models, unit, quantity, unit-price, amount of money, tax, commodity code.

This paper uses Python language to calculate the threshold of each index. First, we import the tables of taxpayer and invoice information:

```python
nsr=pd.read_csv('nsrxx.csv')  # taxpayer info
fp=pd.read.csv('zzsfp.csv')   # invoice content
```

1)Deviation of tax burden rate

The critical value is determined based on the distribution of the deviation.

Code 2. algorithm of deviation
\[
\text{tbr\_mean}=fp\text{.se\_sum()/fp\text{.je\_sum()}}
\]
\[
\text{sum}=\text{grouped[['se','je']].sum()}
\]
\[
\text{tbr\_deviation}=\text{np\_abs(sum\_se/sum\_je/tbr\_mean}-1)
\]
\[
\text{tbr\_deviation}[\text{tbr\_deviation\_isnull()}]=0
\]
\[
\text{plt\_hist(tbr\_deviation,10)}
\]
\[
\text{plt\_show()}
\]

From the distribution map Fig.1 can be seen that 95% of the deviation of enterprises below 0.013. So, the threshold of deviation is 0.013.

2) Fluctuation of tax burden rate

We use the mean of tax burden rate in the window period as the basis for calculating the rate of change.

Code 3. Algorithm of fluctuation

\[
\text{series}=fp\text{.groupby(['xf\_id','kpyf'])[['se','je']].sum()}
\]
\[
\text{series}=\text{series\_se/series\_je}
\]
\[
\text{fluctuation}=\text{np\_abs(series/series\_mean(level=0)}-1)
\]
\[
\text{tbr\_f}=\text{fluctuation\_max(level=0)}
\]
\[
\text{tbr\_f}[\text{tbr\_f\_isnull()}]=0
\]
\[
\text{plt\_hist(tbr\_f,10)}
\]
\[
\text{plt\_show()}
\]
\[
\text{tbr\_f\_quantile(0.95)}
\]

The threshold of fluctuation of tax burden rate is 0.034 which determined by 0.95 quantile as Fig.2.

3) Correlation between tax and income

Normally, the correlation between tax and income growth is positive, and its value is close to 1. If the correlation is negative, the tax revenue and income has appeared the reverse growth. In this case, we have reason to suspect that the enterprise invoice abnormal.

Code 4. Algorithm of correlation

\[
\text{series}=fp\text{.groupby(['xf\_id','kpyf'])[['se','je']].sum()}
\]
\[
\text{series}=\text{series/series\_mean(level=0)}-1
\]
\[
\text{corr}=\text{series\_se/series\_je}
\]
\[
\text{corr}[\text{corr\_isNull()}]=1
\]
\[
\text{corr[np\_isin(corr)]}=1
\]
\[
\text{corr}=\text{corr\_min(level=0)}
\]
\[
\text{corr[corr<0].count()/len(corr)}
\]

By calculating the sample data, we found that there are 48 enterprises with negative correlation between tax and income. The proportion of these enterprises is 1.4, which is about less than that of abnormal enterprises. Therefore, we set the threshold of the correlation indicator to 0.

4) Label of invoice quantity abnormal

We consider the value of the invoice quantity anomaly label from two aspects: the invoice absolute amount as Fig.3 and the invoice relative amount as Fig.4.

Code 5. Distribution chart of invoice absolute amount

\[
\text{series}=fp\text{.groupby(['xf\_id','kpyf'])[['fp\_nid']].count()}
\]
\[
\text{fp\_count}=\text{series\_max(level=0)}
\]
\[
\text{plt\_hist(fp\_count,10)}
\]
\[
\text{plt\_show()}
\]

Code 6. Distribution of relative amount of invoice:

\[
\text{series}=fp\text{.groupby(['xf\_id','kpyf'])[['fp\_nid']].count()}
\]
\[
\text{rate}=\text{series/series\_mean(level=0)}-1
\]
\[
\text{rate\_max}=\text{rate\_max(level=0)}
\]
\[
\text{plt\_hist(rate\_max,10)}
\]
\[
\text{plt\_show()}
\]
By weighted analysis of the two indicators, the critical values for early warning are as follows: if the absolute amount of the current invoice exceeds 20, and the relative amount of the invoice exceeds 200 per cent, then the label is 1, otherwise, the label is 0.

5) Invalid invoice ratio

Invalid invoices include invoices declared invalid for the current period and invoices hedged at a later stage. Because the Red letter invoice does not indicate which month's invoice it is used as hedge against, there will be trouble in calculating invalid invoices. The method of calculating the invalid invoices is to use the amount of invoices in the current period plus the maximum number of red-letter invoices that appear in the window period. Therefore, the calculated invalid invoice ratio may exceed 1.0 as Fig.5, but this does not affect the effectiveness of the indicator.

Code 7. Invalid invoice ratio

```python
fp_cnt = fp.groupby(['xf_id', 'kpyf'])['fp_nid'].count()
invalid = fp[fp.zfbz == 'Y']
invalid_cnt = invalid.groupby(['xf_id', 'kpyf'])['zfbz'].count()
red_letter = fp[fp.se < 0]
red_cnt = red_letter.groupby(['xf_id', 'kpyf'])['fp_nid'].count()
invalid = invalid_cnt.max(level=0) + red_cnt.max(level=0)
invalid[invalid.isnull()] = 0
rate = invalid / fp_cnt.max(level=0)
rate[rate.isnull()] = 0
plt.hist(rate, 10)
plt.show()
rate.quantile(0.95)
```

According to the distribution chart, the critical value of 0.95 quantile is 0.41.

Figure 1. Deviation of tax burden rate. Figure 2. Fluctuation of tax burden rate. Figure 3. Distribution of invoice amount.

Figure 4. Distribution of relative amount of invoice. Figure 5. Distribution of invalid invoice ratio.

**Conclusion**

The man-machine combination method is worth popularizing in determining the tax assessment cases. Because the human-machine combination has both randomness and human’s will, and has stronger adaptability, so it becomes the main method to determine the case-selected of tax assessment. In this paper, five kinds of invoice indicator are extracted from the VAT invoice data, and based on the analysis of a large number of invoice data, the threshold of each index is determined, which provides the basis for the human-machine combination.
References


