Could Replacement of Business Tax with VAT Improve Total Factor Productivity
Empirical Evidence Based on Data Mining and Text Analysis

Hao Li and Siyuan Hong[✉]

SILC Business School, Shanghai University, 20 ChengZhong Road, Shanghai, China
987016010@qq.com

Abstract. In response to the call of "Internet + Taxation", the construction of the Comprehensive VAT Invoice Service Platform has boosted the modernization and informatization of fiscal and taxation management, which played an important role in "the Replacement of Business Tax with Value Added Tax" (RBTVAT) Based on the quasi-natural experimental event of RBTVAT, this paper adopts the text analysis method, using Perl and the "jieba" library in Python to collect and process data, then using the difference-difference model to study the impact of RBTVAT on the total factor productivity (TFP) of Chinese A-share listed companies. The research results show that RBTVAT significantly improves the TFP of enterprises. This paper also confirms that increasing investment in R&D is an impact mechanism of RBTVAT to improve the TFP of enterprises.

Keywords: Tax Reform · Vat · Tfp · R&D Input · Text Analysis

1 Introduction

At a time when information technology construction has become increasingly important, realizing tax informatization to adapt to the Internet era is an important direction for tax system reform [8].

Since the reform and opening up, China's economy has transformed from a high-speed growth stage to a high-quality development stage, which requires the Chinese government to focus on improving the TFP of enterprises, and gradually realize the transformation and upgrading of the economic development model through continuous efficiency improvement. As of 2017, the ratio of the service industry to GDP has exceeded 50%, and it has shown a steady increase year by year. Ping et al. (2017) [6] point out that the TFP of the tertiary sector in China is still lower than that of the secondary sector. If this continues in the long run, the booming tertiary sector may instead lead to the TFP of the overall economy being pulled down. Therefore, it is necessary to carry out corresponding tax reforms to appropriately reduce the tax burden of the tertiary industry. In the process of implementing the RBTVAT from pilot to nationwide, the information interaction between the enterprise side and the government side has significantly influenced the implementation effect of the policy. However, an accurate policy evaluation is inseparable from the support of credible research data. Therefore, the method of text analysis is used

in this article: data collection through Perl, data preprocessing through the "jieba" library, and keyword matching through Python to finally achieve a credible interpretation of the RBTVAT.

2   Theoretical Analysis and Research Hypothesis

VAT overcomes many of the problems of business tax and can play a positive role in the development of enterprises.

Firstly, compared with other turnover taxes, VAT has the advantage of tax neutrality and avoids the double taxation under the business tax system, which in turn can significantly improve the tax burden shifting ability of enterprises [4].

Secondly, VAT is a kind of tax that is not on the price. By reducing the prices of all producers in the production chain, RBTVAT reduces the production costs of enterprises and increases their available cash flow, which lays a material foundation for enterprises to increase their investment in innovation and improve their total factor productivity [3].

In addition, VAT only focuses on the value-added part rather than the overall taxation, which will reduce the motivation of enterprises to adopt an integrated model, promote industrial specialization and division of labor, and exert economies of scale [1, 9].

Therefore, this paper believes that RBTVAT will improve the TFP of enterprises from at least three aspects, and put forward the following assumptions:

H_1: Holding other factors fixed, the TFP of enterprises will be significantly improved after participating in the RBTVAT.

In the neoclassical economic growth model, TFP growth represents technological progress. Empirical evidence from developed countries shows that R&D investment is a key driver of technological progress and an important factor in corporate capital allocation [2]. After the RBTVAT, enterprises reduce their costs and improve the efficiency of resource allocation by separating the auxiliary production business from the core R&D business; and the specialized division of labor upstream and downstream of the industrial chain also promotes the spillover effect of technology as well as the diffusion effect, which eventually leads to the increase of enterprise R&D investment [7].

Therefore, this paper further proposes the following hypothesis:

H_2: RBTVAT improves the TFP of enterprises by increasing the R&D investment of enterprises.

3   Research Design

Given that the RBTVAT is a gradual progress in different regions, industries and time periods, so it can be approximated as a quasi-natural experimental design, which is suitable for difference in difference method (DID) for effect analysis and policy evaluation.

3.1   Sample Selection and Data Sources

The research sample selected for this paper is listed companies in China's A-share market from 2009–2018.

The time when the company joined the RBTVAT is obtained by using Python 3.0 to extract keywords from the company's annual report. The reason for this is that the policy of RBTVAT is a gradual implementation of different regions, industries and time periods. However, previous studies have not identified the specific years in which companies participated in the RBTVAT. Therefore, this paper adopts the method of Text Analysis, using Perl web page capture algorithm programming, we downloaded the annual reports of all listed companies from Cninfo.com, which is the information disclosure website of listed companies designated by China Securities Regulatory Commission, and then uses the "jieba" library in Python to extract "ying ye shui gai zeng zhi shui" (replacing business tax with value-added tax) or "ying gai zeng" (RBTVAT) as the keywords to accurately identify the specific time for listed companies to implement the change from business tax to value-added tax.

The enterprises' financial data is obtained from the China Stock Market & Accounting Research Database (CSMAR).

In this paper, the initial samples are further processed by using Stata 16.0: (1) excluding samples that are cross-listed in B-shares or H-shares; (2) excluding samples with ST, ST* or delisted; (3) excluding samples with missing data on study variables; (4) excluding companies with abnormal data and asset-liability ratios greater than 1; (5) in order to exclude interference from extreme outliers, all continuous variables are Winsorized at 1% and 99% percentile.

3.2 Variable Definitions

The explained variables variable is TFP calculated by the Solow residual method; the core explanatory variable is tax_after. When an enterprise participates in RBTVAT, the value tax_after is defined as 1 for both the current year and the year after participation, otherwise the value of tax_after is 0. The mediating variable is the level of R&D input. In addition, in order to maximize the explanatory power, a series of control variables are added in this paper. All variables are summarized in Table 1.

3.3 Empirical Model Building

In order to verify H_1, this paper constructs the following difference-in-difference model to test the relationship between RBTVAT and firms' TFP:

$$
\begin{aligned}
tfp\_ols = {} & \alpha_0 + \alpha_1 tax\_after + \alpha_n Controls \\
& + Firm + Year + \varepsilon
\end{aligned}
\tag{1}
$$

In formula (1), the coefficient of tax_after measures the impact of the policy of replacing business tax with value-added tax on the total factor productivity of enterprises. If H_1 holds that the total factor productivity of firms will increase after participating in the policy of RBTVAT, then α_1 should be significantly positive.

To verify the role of the RBTVAT in improving the R&D investment of enterprises, this paper adds the enterprise R&D investment "rdtax" after the RBTVAT into the formula

Table 1. VARIABLE DEFINITION.

Variable	Definition
tfp_ols	TFP calculated by the Solow residual method
tfp_op	TFP calculated by the OP method
tfp_lp	TFP calculated by the LP method
tax_after	In the current year, whether the company participated in the RBTVAT, 1 means participating, 0 means not participating
rdtax	R&D investment/total assets
size	Natural logarithm of the company's total assets at the end of the period
roa	Net profit after tax/total assets
tobinq	Market value of the company/replacement cost of assets
leverage	Total liabilities/total assets
cashratio	(Money funds + marketable securities) / current liabilities
cashflowdebtratio	Total net cash flow from operating activities/total debt
ocfcapital	(Gross cash flow from operations - cash flow available to the firm after the increase in operating capital) / capital
cashcapital	(Cash + cash equivalents)/capital
dividendcapital	Dividends/Capital
subsidy	Government subsidy income/main business income
firm	Individual fixed effects
year	Annual fixed effects

(1), and constructs a formula (2), to explore the mediating effect of corporate R&D investment on the relationship between RBTVAT and corporate's TFP.

$$tfp\_ols = \alpha_0 + \alpha_1 tax_{after} + \alpha_2 rdtax + \alpha_n Controls \\ + Firm + Year + \varepsilon \tag{2}$$

If the RBTVAT does indeed increase the total factor productivity by increasing the R&D investment of enterprises, then α_2 should be significantly positive.

4   Empirical Analyses

To test H_1, the mainline regression results are shown in Table 2. Columns (1) is the regression results without control variables. Columns (2) is the regression result of adding control variable. It is apparent that tfp_ols and tax_after show a significant positive relationship with or without the addition of control variables.

As mentioned above, the implementation of RBTVAT has increased the R&D investment of enterprises, and there is a positive correlation between R&D investment and

Table 2. REGRESSION RESULTS

	(1) tfp_ols	(2) tfp_ols
tax_after	0.0196**	0.0170*
	(2.02)	(1.81)
Control	No	Yes
N	26721	26005
adj. R2	0.397	0.433
Firm Fixed Effect	YES	YES
Year Fixed Effect	YES	YES

Note: T values in brackets, *, * *, and * * * indicate significant differences at 10%, 5%, and 1% significance levels, respectively.

Table 3. REGRESSION RESULTS OF IMPACT MECHANISM.

	(1) tfp_ols
tax_after	0.0389***
	(4.13)
rdtax	0.558***
	(3.05)
Control	Yes
N	26005
adj. R2	0.440
Firm Fixed Effect	YES
Year Fixed Effect	YES

Note: T values in brackets, *, * *, and * * * indicate significant differences at 10%, 5%, and 1% significance levels, respectively.

total factor productivity [5]. Therefore, this paper examines the impact mechanism of the RBTVAT on the TFP of enterprises, and the results are shown in Table 3.

It can be observed that the regression coefficient of enterprise R&D investment is significant at the 1% level, which means that enterprise R&D investment has a partial mediating effect on the relationship between RBTVAT and enterprise's TFP, which supports H_2.

However, to satisfy the prerequisites for the use of the DID method, the parallel trend test was conducted in this paper. Table 4 show that there is no significant change

Table 4. PARALLEL TREND TEST.

	(1) tfp_ols
Before6	0.0191
	(1.11)
Before5	0.0140
	(0.71)
Before4	0.0184
	(0.90)
Before3	0.0309
	(1.47)
Before2	0.0242
	(1.15)
Before1	0.0312
	(1.45)
Current	0.0402*
	(1.87)
After1	0.0332
	(1.49)
After2	0.0369*
	(1.65)
After3	0.0822**
	(2.57)
After4	0.0621*
	(1.78)
After5	0.0663*
	(1.77)
After6	0.112**
	(2.40)
N	26005
adj. R2	0.433
Control	YES
Firm Fixed Effect	YES
Year Fixed Effect	YES

Note: T values in brackets, *, * *, and * * * indicate significant differences at 10%, 5%, and 1% significance levels, respectively.

Table 5. PSM-DID.

| | (1) |
	tfp_ols
tax_after	0.0170*
	(1.81)
Control	Yes
N	26005
adj. R2	0.433
Firm Fixed Effect	YES
Year Fixed Effect	YES

Note: T values in brackets, *, * *, and * * * indicate significant differences at 10%, 5%, and 1% significance levels, respectively.

in the TFP of enterprises before participating in the RBTVAT, while two years after participating, the TFP of enterprises increased significantly in each year, thus verifying the robustness of the mainline regression results.

In order to effectively evaluate the real effect of the policy, this paper uses the PSM-DID research method to conduct year-by-year matching to find a suitable control group for the treatment group of that year. The PSM method can effectively control the influence of characteristic differences between the enterprises. Thereby further reducing the accidental differences between enterprises and enhancing the comparability between individuals.

The regression results of the PSM-DID model are shown in Table 5. After controlling for the front side of DID using PSM, the regression coefficients of tax_after are still significantly positive at the 10% level. After suppressing the pre-existing differences among firms, the RBTVAT still has a significant contribution to the TFP of firms, which also enhances the robustness of H_1.

To further enhance robustness, this paper replaces the former explained variables, which is calculated using the Solow residual method. The total factor productivity of the firm is recalculated using the OP method and the LP method, and added to the regression as new explained variables respectively. As shown in Table 6, after re-regression, the coefficients of tax_after are still significantly positive at the 1% and 10% confidence levels, respectively.

On this basis, this paper performs PSM-DID regression again on the model after replacing the explained variables. The regression results are shown in Table 7.

When the explained variable is the total factor productivity measured by the OP method, the coefficient of tax_after is significantly positive at the 1% confidence interval; when the explained variable is the total factor productivity measured by the LP method, the coefficient of tax_after is significantly positive at the 10% confidence interval. The robustness of this paper becomes stronger.

Table 6. SBUSTITUTION OF EXPLANATORY VAVIABLE.

	(1) tfp_op	(2) tfp_lp
tax_after	0.0602***	0.0492*
	(7.57)	(1.75)
Control	Yes	Yes
N	25881	26009
adj. R2	0.642	0.889
Firm Fixed Effect	YES	YES
Year Fixed Effect	YES	YES

Note: T values in brackets, *, * *, and * * * indicate significant differences at 10%, 5%, and 1% significance levels, respectively.

Table 7. SBUSTITUTION OF EXPLANATORY VAVIABLE (PSM-DID).

	(1) tfp_op	(2) tfp_lp
tax_after	0.0790***	0.0171*
	(7.37)	(1.73)
rdtax	0.481**	0.340*
	(2.27)	(1.68)
Control	Yes	Yes
N	25881	26009
adj. R2	0.643	0.892
Firm Fixed Effect	YES	YES
Year Fixed Effect	YES	YES

Note: T values in brackets, *, * *, and * * * indicate significant differences at 10%, 5%, and 1% significance levels, respectively.

5 Conclusions

This paper takes the RBTVAT as the entry point, and use Perl, "jieba" library and Python to achieve the accurate processing of data, and then explore the impact of the tax system reform on the TFP of enterprises in this information era.

The empirical results show that RBTVAT can effectively improve the TFP of enterprises, and the R&D investment plays a positive mediating role in the relationship

between the two. Based on the above findings, in order to better promote the standardization and informatization of the value-added tax system, this paper proposes the following policy recommendations for goverment and enterprises:

Based on the positive intermediary role of enterprises' R&D investment, the government should improve the construction of tax information platform. Specifically, firstly, the government should promote the cooperation between the State Taxation Bureau and the local taxation bureaus, deepen the interconnection, integration and utilization of information resources. Secondly, the government should also strengthen the cooperation with social and international organizations to expand the scope and enrich the content of data that can be shared. While further improving the VAT deduction chain in China through the above measures, the government should also expand the scope of deduction for R&D activities and increase the deductible items of R&D expenses, etc., to increase the enthusiasm of enterprises for R&D.

Based on the positive effect of RBTVAT on TFP, enterprises should carry out informatization transformation of their own tax management systems. While RBTVAT brings policy dividends to enterprises, it also poses challenges to a series of fiscal and taxation management practices such as invoice management and VAT input tax certification, which requires enterprises to comprehensively improve their own fiscal and taxation management informatization level, striving to modernize taxation. The informationization of financial management system will enable enterprises to enjoy the benefits of policies to the greatest extent, reduce their tax expenditures and production costs, and enable enterprises to invest this part of the new cash flow in R&D activities.

References

1. Chen Z., Wang Y., 2016, "Does VAT Reform promote division of labor: Evidence from Chinese listed companies", Management World, (03), pp. 36–45+59.
2. Griliches Z., 1986, "Productivity, R&D and Basic Research at the Firm Level in the 1970s", American Economic Review, 76(1), pp. 141–154.
3. Liu BH., 2015, "Measurement and Prediction of Price Effects of VAT Reform - An Analysis Based on Input-Output Tables", Finance & Trade Economics, (10), pp. 59–72.
4. Liu BH., Kou EH. and Yang LJ., 2019, "Multiple VAT TaxRates, Resource Misallocation and Total Factor Production Loss", Economic Research Journal, 54(05), pp. 113–128.
5. Peng F., Mao DF., 2018, "Does expensive paddy and cheap rice restrain the misallocation in rice processing industry? Based on the micro data of Chinese firms", Industrial Economics Research, (01), pp. 52–64+89.
6. Ping XQ., Huang X. and An R., 2017, "The TFP Heterogeneity and Tax Burden Difference between China's Services Industry and Manufacturing Industry: Pre-Transformation from Business Tax to VAT", Comparative Economic & Social Systems, (02), pp. 77–84.
7. Qian XD., 2018, "Replacing Business Tax with VAT, Specialized Division and R&D Investment——Empirical Study Based on Micro-Enterprise Data", Journal of Shanxi University of Finance and Economics, 40(02), pp. 98–111.
8. Song YF., 2016, "Information construction to help the transformation of camp reform to the ground", China State Finance, 2016(17), pp. 67
9. Wu SM., Li ST., 2013, "Economic Complexity and its Evolution in China: Based on Input-Output Tables from 1987 to 2007", Management Review, 25(03), pp. 3–11+17.