Finance and Trade Dynamics
WISDOM ACADEMIC https://doi.org/10.63808/ftd.v2i3.487 Volume 2 , Issue 3

Research Article

Big Data Tax Enforcement and Corporate Earnings Management — Evidence from the Golden Tax Project Phase IV

Guangwei Hao[1,*]
[1]Central University of Finance and Economics, Beijing 100081, China
[*]**Correspondence:** Guangwei Hao, 174238223@qq.com

Abstract

An issue under examination here is whether data-driven tax administration leads to governance externalities in financial reporting. In this paper, I use the staggered introduction of China's fully digitalized e-invoice issuance pilot, the dated element of the fourth phase of the Golden Tax project. Using 23,256 firm-year observations of China's listed A-share firms from 2018 to 2024, I find that a firm-year's participation in the pilot leads to a reduction in absolute performance-matched discretionary accruals by 0.0050, around 8.2 per cent of the sample mean, and a reduction in real earnings management with an insignificant effect size. The equality of the standardized effects cannot be rejected at the five percent significance level against the signed measure of real manipulation, but it can be marginally rejected against its absolute counterpart, and therefore the asymmetric effect is qualified. Reductions in analyst forecast error and discretionary book-tax difference are consistent with information and tax-avoidance channels, respectively, and the effect occurs mainly for non-state-owned firms, firms with low-quality audits, and input-intensive industries. Inference at the treatment assignment level results in a wild cluster bootstrap p-value of 0.058.

Keywords

Fully digitalized e-invoice, Gig data tax enforcement, Earnings management, Staggered difference-in-differences, Information transparency

1. Introduction

Tax administration is moving away from traditional ex post audits based on sampling towards transaction-based monitoring with real-time e-invoicing and third-party validation. Such monitoring serves as a consequential monitor that reduces resources available to outside investors. As tax and financial reporting use the same transaction data, enforcing the former

limits flexibility in the latter, with research indicating that monitoring improves reporting quality and that e-invoicing increases both revenues and reported taxes. It is unknown whether monitoring through continuous data analysis generates spillovers of a different nature. The design of China's Golden Tax Project Phase IV (GTP IV) is described in March 2021 Opinions on Deepening Tax Collection Reform. The key feature of the project is a completely digitized electronic invoice to be used in a nation-wide unified online platform available to the tax authority in real time. Being an ongoing program, treatment date is defined by the start of pilot implementation in a firm's registration jurisdiction. The invoice does not involve stock or firm-specific quota and goes directly to the tax authority at issuance, thus moving enforcement from invoices to data and making it possible to cross-check the reported transaction chain against the counterpart's transaction history.

Evidence of firm-level effects is available only for GTP III, where the project decreased tax avoidance, audit fees, increased tax compliance, and reduced debt default risk, but there is little evidence on GTP IV, which is linked to green innovation. The difference is that GTP III compares the declaration with the firm's own files, whereas GTP IV can validate transaction chain based on the counterparty's files. Identification exploits the staggered rollout of the pilot program among 36 provincial tax jurisdictions between December 2021 and December 2023 and nationwide implementation on 1 December 2024. Since entry sequence is based on administrative preparedness rather than characteristics of firms, treatment timing is arguably exogenous to accruals (test rather than assume).

Based on 23,256 firm-year observations spanning 2018-2024, being assigned to the pilot program decreases absolute, performance-matched discretionary accruals by 0.0050 (approximately 8.2% of the mean of the sample); the effect on real earnings management is statistically indistinguishable from zero. It transforms the nature of governance externalities of tax enforcement from documentary to data systems and reveals an asymmetry in the manipulation technologies: government-based enforcement cannot substitute private control, as the former captures actions but not their legitimacy. There is no “never treated” group left; p-value for the wild cluster bootstrap is 0.058.

2. Literature Review and Hypothesis Development

2.1. Tax enforcement and its governance spillovers

Tax collection brings not only financial advantages since concealment of income and shifting of resources away from foreign investors involve similar complicated transactions. As tax enforcement is effective, the cost of opaqueness increases and the amount of private-benefit extraction decreases, evidenced by correlation between monitoring/reporting

quality and reduction of default risk under digital governance. This correlation is not unconditional because avoidance and active reporting are managed by managerial incentives, which means that as the cost of one rises, the effort is transferred to another one.

A variety of papers exploit the third phase as a natural experiment to investigate tax evasion and position reporting. Similarity between these works lies in the informational channel: digital tax collection enhances transparency of reporting, forcing auditors to decrease fees.

There are two shortcomings that provide motivation to conduct this research. Firstly, in the third phase, the observation was based on declarations of invoices, while digitalization makes the unit coincide with the transaction event itself with the counterpart. Secondly, earnings management is considered as one entity, although these techniques are different in whether they produce a documentable trail.

2.2. Hypothesis development

Accrual-based earnings management involves manipulation of the timing, recognition, and estimation without any change in the economic reality. In China, there are several techniques like early revenue recognition, channel stuffing, and inappropriate cost capitalization that result in the creation of documents that do not conform to other accounting documents filed by the firm later on. The digital invoicing system makes it possible to spot such discrepancies instantly.

H1. Entry into the fully digitalized e-invoice issuing pilot is associated with a reduction in accrual-based earnings management.

However, the forecast regarding real earnings management is an indirect one. Real manipulation varies according to the time real transactions happen, not according to how they are recorded, thereby generating an invoice that matches the accounting record of the counterpart, and appears odd only relative to fundamentals, which are not easy for the tax authority to evaluate. The theory of substitution posits that real manipulation is the result of accrual manipulation becoming expensive.

H2. The restraining effect on real earnings management is weaker than the restraining effect on accrual-based earnings management.

3. Research Design

3.1. Sample selection and data

This study uses Shanghai and Shenzhen listed firms in 2018-2024, using winsorised continuous variables at the 1st and 99th percentiles. Data on firms are obtained from the

WISDOM ACADEMIC https://doi.org/10.63808/ftd.v2i3.487 Volume 2 , Issue 3

China Stock Market and Accounting Research database [14], while the launching dates for each jurisdiction were collected manually from announcements by the State Taxation Administration and provincial bureaux, as listed in Table A2. Out of 32,009 firm years of A-share firms, 1,043 are financial firms, 1,586 are treated under special treatment, 312 relocate, and 5,812 have missing data or insufficient lags. This results in 23,256 observations among 4,185 firms. Table 1 provides the distribution and balance of cohorts.

The pilot is implemented by 36 provincial level tax bureaux (31 provincial units plus Dalian, Ningbo, Xiamen, Qingdao, and Shenzhen). Shenzhen is not part of the initial Guangdong province pilot program. Each firm was assigned according to the jurisdiction that its registered office is located in during 2018, because listed firms run subsidiaries in multiple jurisdictions.

Table 1

Treatment cohorts and cohort balance

Panel A. Treatment cohorts	Jurisdictions	Firms	Firm-years	Launch window
Onset 2022	3	689	3,954	1 Dec 2021
Onset 2023	21	2,288	12,766	Oct 2022 - Apr 2023
Onset 2024	12	1,208	6,536	Oct 2023 - Dec 2023
Total	36	4,185	23,256	

Panel B. Cohort balance on pre-treatment characteristics, 2018-2021

Variable	Onset 2022	Onset 2023	Onset 2024	p-value
Size	22.36	22.33	22.58	0.002
Lev	0.4062	0.4074	0.4231	0.014
ROA	0.0384	0.0371	0.0331	0.041
MB	3.1284	3.0417	2.8735	0.006
Top1	0.3187	0.3241	0.3372	0.008
SOE	0.2914	0.3082	0.4046	<0.001
|DA|	0.0619	0.0616	0.0603	0.412
Annual change in |DA|, 2018-2021	0.0004	0.0002	0.0003	0.638

Note. In Panel A, firms are assigned to the registered office jurisdiction for 2018 by the

halfway rule. By 2024, all jurisdictions have been covered, and therefore there is no never-treated group. In Panel B, pre-treatment means for each cohort are provided together with p-values from equality tests conducted jointly for each jurisdiction. The level differential is explained by the rollout geographies and is captured by firm fixed-effects; all we need is identical levels and trends of |DA|.

3.2. Variable measurement

Accrual-based earnings management is measured by the absolute value of performance-matched discretionary accruals, estimated by industry and year from equation (1) with a minimum of ten observations per industry-year [11].

$$\frac{TA_{i,t}}{A_{i,t-1}}=\alpha_1\frac{1}{A_{i,t-1}}+\alpha_2\frac{\Delta REV_{i,t}-\Delta REC_{i,t}}{A_{i,t-1}}+\alpha_3\frac{PPE_{i,t}}{A_{i,t-1}}+\alpha_4 ROA_{i,t-1}+\varepsilon_{i,t} \quad (1)$$

TA denotes total accruals, A lagged total assets, ΔREV and ΔREC the change in sales and receivables, PPE gross property, plant and equipment, and ROA lagged return on assets, with the absolute residual denoted |DA|. Real earnings management combines abnormal production costs, operating cash flow and discretionary expenditure as in equation (2) [12]. Because |DA| is unsigned while REM is signed, |REM| is reported alongside REM wherever the two are compared.

$$REM_{i,t}=ABPROD_{i,t}-ABCFO_{i,t}-ABDISX_{i,t} \quad (2)$$

The implementation occurred on two margins: receipt date for the former and issuance date for the latter, which determines the timing of treatment. Receipt, where individuals may claim invoices issued outside their province, became available to all 36 jurisdictions on 28 August 2022, whereas issuance was introduced gradually between December 2021 and December 2023. Only the issuance margin affects the invoice issuance process, and thus GTPIV is the year that the jurisdiction starts issuing invoices via the pilot program. Given the early expanding receipt margin, not-yet-treated firms are partially affected, biasing estimates toward zero. Onset is the year when the treatment is introduced up to 30 June; otherwise, the next year. Thus, the cohort assignment does not depend on the exact day within a half-year period. An announcement introduces the pilot program to the subset of taxpayers identified by the provincial bureau, and jurisdictional onset constitutes intention-to-treat allocation.

Controls include firm size, leverage, profitability, sales growth rate, operating cash flow, ownership concentration, board independence, CEO duality, auditor identity, market-to-book, and a loss indicator. Listing age is controlled for as a function of the calendar year deterministically. Information opacity is captured by absolute analyst forecast error and forecast dispersion (without using discretionary accruals). Tax avoidance is operationalized as book-tax gap and its discretionary portion—the residual from regressing the former on total

accruals [2]. Input intensity is defined as the median ratio of operating costs to operating income in two-digit industry groups. Table A1 describes all variables and their sources.

3.3. Identification strategy

The baseline is equation (3), in which EM denotes |DA|, REM or |REM|.

$$EM_{i,t}=\alpha+\beta\, GTPIV_{i,t}+\gamma' Controls_{i,t}+\mu_i+\lambda_t+\varepsilon_{i,t} \quad (3)$$

Firm fixed effects μ absorb time-invariant jurisdiction and firm characteristics, so identification derives only from onset timing, and λ denotes year fixed effects. Parallel trends are assessed with equation (4), in which the year preceding onset is the omitted reference and the earliest lead is binned.

$$EM_{i,t}=\alpha+\sum_{k=-4,\ k\neq-1}^{+1}\delta_k D_{i,t}^{k}+\gamma' Controls_{i,t}+\mu_i+\lambda_t+\varepsilon_{i,t} \quad (4)$$

There are three considerations. First, since all jurisdictions will be treated by 2024, there is no never-treated group and there is no variability in the treatment status in 2024. The identification strategy uses 2022 (three treated, thirty-three never) and 2023 (twenty-four treated, twelve never). We provide event-study estimates up until one year post-onset plus a specification from 2018 to 2023 where twelve jurisdictions are the treatment group. Staggered adoption and heterogeneous effects mean that the two-way fixed effect estimator will be a very poor average of group-time effects, hence we re-estimate using the group-time estimator, the interaction-weighted estimator, and the imputation estimator.

The treatment is jurisdiction-specific but the clusters are firm-specific resulting in underestimated standard errors for the estimation across the 36 jurisdictions. The jurisdiction-cluster wild bootstrap is the conservative test and event study plots show the confidence interval. The non-rejection of the pre-trends cannot be taken as absence of pre-trends since power is very small.

4. Empirical Results

4.1. Baseline regression results

The results of the baseline estimations are shown in Table 2. The estimates in columns (1)-(2) are based on accrual-based earnings management; both coefficients of GTPIV are negative and significant at 1% level of significance. An estimate of -0.0050 represents around 8.2% of the sample mean of |DA| and around 0.085 of its SD, which supports H1. Its small magnitude corresponds to the design intention-to-treat, and exposure is estimated via the registered office.

The results of the estimation of real earnings management coefficients are given in columns (3)-(5); both estimates are insignificant. As the variables do not have the same scale,

Panel B standardizes all the series and tests H2 within the framework of the seemingly unrelated equations model. Rejection of equality can be observed for the signed version (chi-square 4.64, p=0.031), and marginally for the absolute measure (chi-square 3.34, p = 0.068). Some asymmetry can be explained by an unsigned nature of |DA| compared to the signed nature of REM. Thus, H2 is partially supported. Its point estimate is negative, and a substitution interpretation of it does not find any support.

Table 2

Baseline regression results

Panel A. Baseline estimates

	(1)	(2)	(3)	(4)	(5)
Dependent variable	**|DA|**	**|DA|**	**REM**	**REM**	**|REM|**
GTPIV	-0.0061***	-0.0050***	-0.0068	-0.0041	-0.0038
	(0.0018)	(0.0017)	(0.0043)	(0.0040)	(0.0031)
Controls	No	Yes	No	Yes	Yes
Firm fixed effects	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes
Observations	23,256	23,256	23,256	23,256	23,256
Adjusted R-squared	0.216	0.247	0.398	0.436	0.402

Panel B. Standardised effects and tests of equality

	Standardised |DA|	Standardised REM	Standardised |REM|	Chi-square	p-value
GTPIV	-0.085***	-0.019	-0.027		
	(0.0289)	(0.0184)	(0.0222)		
Equality against REM				4.64	0.031
Equality against |REM|				3.34	0.068

Note. Controls: Size, Lev, ROA, Growth, CFO, Top1, Indep, Dual, Big4, MB, Loss (Table A1; available upon request). Panel B rescales each dependent variable to unit variance

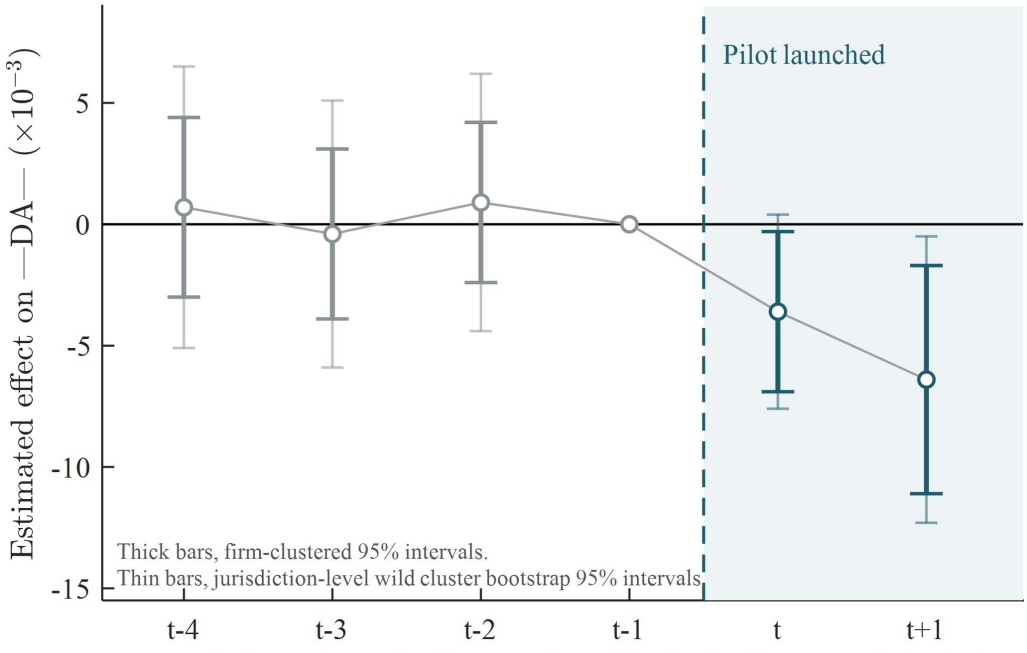
without altering t-statistics and uses a joint test for equality based on a seemingly unrelated system, thus chi-square is larger than the individual equation chi-square since the residual correlation is positive. Firm-clustered SEs (in parentheses); jurisdiction-clustered testing (Table 5). *** p<0.01, ** p<0.05, * p<0

4.2. Parallel trends and dynamic effects

Figure 1 presents the coefficients from equation (4), four years prior to the onset through the year following the onset, binned into earlier leads. There is no significant effect among the coefficients prior to the intervention period and there is no clear trend observed despite the wide confidence interval bounds that indicate a pre-trend cannot be ruled out, that is, absence of evidence for a pre-trend, not evidence of a pre-trend [20]. The effect is negative in the year of onset and greater the year after.

Figure 1

Dynamic effects of the pilot launch on accrual-based earnings management



Notes. Coefficients from equation (4) on |DA|, in units of 0.001. The year before launch is the omitted reference and the earliest lead is binned. Thick bars are firm-clustered 95 per cent intervals, thin bars jurisdiction-level wild cluster bootstrap intervals. Coefficients beyond one year are omitted because every jurisdiction is treated by 2024, leaving only already-treated comparisons.

The two sets of intervals give the methodology a lesson. With firm-level clustering, the onset-year effect is significant at the 5% level, while with jurisdiction-level clustering only the year-after-onset effect is significant. The difference arises because each event-time effect

is based on very few treated jurisdictions.

Panel B (Table 1) contrasts the pre-treatment characteristics of the cohorts. Unbalanced levels: The later-entry jurisdictions are more likely to be state owned because of their rollout geography, not reporting behavior. There is no difference between the levels or pre-treatment trends in |DA| between the cohorts.

4.3. Robustness checks

Table 3 has a total of thirteen tests; the test result is informative compared to a uniformly significant table. The period restriction to 2018–2023 results in -0.0049, where the heterogeneity-robust estimators [16], [17], [18] fall between -0.0050 and -0.0057; thus, the negative weight is not causing the effect. With entropy balancing and the removal of the first three jurisdictions, the estimate stands.

The three tests that diminish the estimate are informative. Clustering on the jurisdiction level results in a wild cluster bootstrap p-value of 0.058; thus, the estimate becomes marginal since the analysis level matches the treatment assignment level. Omitting the years 2020 and 2022 weakens the estimate to a 10% significance level since the omission of 2022 leaves only one year of treatment for the earliest jurisdiction; the calendar year coding for onset does this too. The placebo assigning the onset to random results in 0.0001 on average; the true estimate is at the 1.2th percentile.

Table 3

Robustness checks

Specification	GTPIV	Std. error	Observations
(1) Window restricted to 2018-2023	-0.0049**	(0.0019)	19,521
(2) Group-time ATT, not-yet-treated comparisons [16]	-0.0057***	(0.0021)	23,256
(3) Interaction-weighted estimator [17]	-0.0050***	(0.0018)	23,256
(4) Imputation estimator [18]	-0.0054***	(0.0019)	23,256
(5) Entropy balancing on pre-treatment covariates	-0.0055***	(0.0018)	23,256
(6) Excluding the three 2022-cohort jurisdictions	-0.0044**	(0.0020)	19,302
(7) Discretionary accruals without performance matching	-0.0046***	(0.0016)	23,256
(8) Income-increasing accruals only	-0.0061***	(0.0023)	12,338
(9) Controlling for concurrent policies	-0.0048***	(0.0016)	23,256

(10) Excluding 2020 and 2022	-0.0038*	(0.0021)	16,480
(11) Alternative onset convention	-0.0035*	(0.0020)	23,256
(12) Jurisdiction clusters, wild cluster bootstrap [19]	-0.0050*	[p = 0.058]	23,256
(13) Placebo, 1,000 random reassignments	0.0001	[0.0019]	23,256

Note. Each row is a separate estimation on |DA| with the full control set and firm and year fixed effects, and rows (1) to (11) cluster by firm. Row (9) adds indicators for the 2022 value-added tax credit refund programme and the registration-based IPO reform. Row (12) gives the wild cluster bootstrap p-value over thirty-six clusters, and row (13) the mean across 1,000 simulations with the simulated standard deviation in brackets. *** p<0.01, ** p<0.05, * p<0.10.

5. Further Analysis

5.1. Mechanism tests

Two mechanisms were suggested: reduced information opacity and decreased tax avoidance space. Panel A of Table 4 indicates that absolute forecast errors and discretionary book-tax differences become significant post-treatment whereas forecast differences decline insignificantly and raw book-tax differences become insignificant after treatment indicating weaker information effects. The proxies for information effects do not use discretionary accruals and discretionary book-tax difference controls for accrual effects whereas raw book-tax difference does not control for accruals.

Panel B investigates whether the effect becomes significant only where there is greater space for channels to work. The sample was split based on median forecast dispersion in pre-treatment period. Only high dispersion subsample indicated a significant effect. The same is found for pre-treatment discretionary book-tax differences. These findings confirm rather than identify because splits cannot exclude correlation of firm characteristics.

Table 4

Mechanism tests

Panel A. Effect of the reform on the proposed channels				
Dependent variable	(1) ANAFERR	(2) FDISP	(3) DDBTD	(4) BTD
GTPIV	-0.0026**	-0.0011	-0.0021***	-0.0018*
	(0.0013)	(0.0007)	(0.0008)	(0.0011)

WISDOM ACADEMIC https://doi.org/10.63808/ftd.v2i3.487 Volume 2 , Issue 3

Controls and fixed effects	Yes	Yes	Yes	Yes
Observations	19,842	17,106	22,318	22,318
Adjusted R-squared	0.312	0.284	0.196	0.213

Panel B. Effect on |DA| across channel-based subsamples

Pre-treatment split variable	High	Low	Difference	p-value
Analyst forecast dispersion	-0.0081***	-0.0019	-0.0062	0.028
	(-2.94)	(-0.78)		
	N = 7,842 and 7,769			
Discretionary book-tax difference	-0.0077***	-0.0026	-0.0051	0.049
	(-2.86)	(-1.02)		
	N = 10,468 and 10,394			

Note. Panel A estimates equation (3) with the channel variables as dependents, defined in Table A1. Panel B splits at the industry-year median of the pre-treatment characteristic, t-statistics in parentheses, between-group p-values from 1,000 bootstrap replications. *** p<0.01, ** p<0.05, * p<0.10.

5.2. Heterogeneity analysis

The effect of reform will be the largest when manipulation costs rise, under conditions of lowest constraint and highest exposure to it. Table 5 illustrates the results obtained through three partitions. Ownership reveals that reforms have an effect on non-state-owned firms, but not state-owned firms, as the latter are under more scrutiny (the difference is marginally significant). Audit quality shows that the effect is present only under low audit quality, suggesting substitution between public enforcement and private assurance mechanisms. Finally, the effect is large for invoice-intensive industries.

Table 5

Heterogeneity analysis

Partition	Subsample 1	Subsample 2	Difference	p-value
Panel A. Ownership	Non-state-owned	State-owned		
GTPIV	-0.0068***	-0.0021	-0.0047	0.056
	(-3.05)	(-0.79)		
Observations	15,544	7,712		
Panel B. Audit quality	Low	High		
GTPIV	-0.0067***	-0.0018	-0.0049	0.048
	(-3.12)	(-0.61)		
Observations	15,805	7,451		
Panel C. Input intensity	High	Low		
GTPIV	-0.0079***	-0.0022	-0.0057	0.031
	(-3.28)	(-0.87)		
Observations	12,061	11,195		

Note. The dependent variable is |DA|. Subsamples are estimated separately with the full control set and firm and year fixed effects, and observations sum to 23,256 per panel. SOE, HighAudit and InpInt are defined in Table A1. Firm-clustered t-statistics in parentheses, between-group p-values from 1,000 bootstrap replications. *** p<0.01, ** p<0.05, * p<0.10.

6. Conclusions

Building on the staggered adoption of China's 36 province-wide trials of e-invoices, the evidence indicates that onset reduces earnings management through accruals by about 8% of the sample mean, particularly one year after onset; real earnings management remains about the same. Tax avoidance mechanisms receive more empirical support than information mechanisms, and the effects are pronounced among non-state-owned firms, firms with lower audit quality, and input-intensive industries. The takeaway message is that data-driven compliance enforcement depends on the limitations of available data: transaction-level information detects fabricated or mis-timed transactions but does not convey whether a transaction is really a transaction. The regime improves the quality of accounting reporting without improving the quality of the underlying transaction, so accrual quality does not necessarily reflect decision-making quality. Data on tax may assist in risk-based review

sampling, and the role of auditor skepticism about business decisions increases. There are five limitations, two of which are significant. First, inference at the treatment level is based on 36 clusters, and a wild cluster bootstrap p-value of 0.058 makes the result statistically insignificant, although further replication will become possible with longer panel data. Second, the rollout lacks a never-treated control group, and therefore identification is based on never-treated groups for two treatment groups, with the identification horizon stretching one year after onset. Third, GTP IV is enforced solely through its invoice digitalization module, and the pilot project applies to only a subset of taxpayers based on office registration, which generates measurement error. Fourth, the analysis focuses on listed firms, while unlisted firms are more likely to be affected.

Acknowledgments

The authors declare that no specific acknowledgments are required.

Author contributions

Not applicable.

Funding

This research received no external funding.

Data availability

No datasets were generated or analysed during the current study.

Declarations

Conflict of interest

The authors declare that there is no conflict of interest.

Ethics approval

Not applicable.

Consent to participate

Not applicable.

Consent for publication

All authors have given their consent.

Publisher's Note

Wisdom Academic Press Ltd. remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

[1] Desai, M. A., Dyck, A., & Zingales, L. (2007). Theft and taxes. *Journal of Financial Economics*, *84*(3), 591–623. https://doi.org/10.1016/j.jfineco.2006.05.005

[2] Desai, M. A., & Dharmapala, D. (2006). Corporate tax avoidance and high-powered incentives. *Journal of Financial Economics*, *79*(1), 145–179. https://doi.org/10.1016/j.jfineco.2005.02.002

[3] Hanlon, M., Hoopes, J. L., & Shroff, N. (2014). The effect of tax authority monitoring and enforcement on financial reporting quality. *The Journal of the American Taxation Association*, *36*(2), 137–170. https://doi.org/10.2308/atax-50820

[4] Bellon, M., Dabla-Norris, E., Khalid, S., & Lima, F. (2022). Digitalization to improve tax compliance: Evidence from VAT e-invoicing in Peru. *Journal of Public Economics*, *210*, 104661. https://doi.org/10.1016/j.jpubeco.2022.104661

[5] Li, J., Wang, X., & Wu, Y. (2020). Can government improve tax compliance by adopting advanced information technology? Evidence from the Golden Tax Project III in China. *Economic Modelling*, *93*, 384–397. https://doi.org/10.1016/j.econmod.2020.08.009

[6] Su, L. (2025). Tax administration digitization and corporate tax avoidance: A quasi-natural experiment based on the Golden Tax III project in China. *Baltic Journal of Economics*, *25*(1), 175–199. https://doi.org/10.1080/1406099X.2025.2488723

[7] Wu, X., Ying, S. X., You, J., Wu, X., & Wu, H. (2026). The effect of big data-driven tax enforcement on audit pricing: Evidence from China. *The British Accounting Review*, *58*(1), 101531. https://doi.org/10.1016/j.bar.2024.101531

[8] Chen, W., & Xu, J. (2024). Can digital tax enforcement reduce the risk of corporate debt default? *Economic Analysis and Policy*, *83*, 1041–1060. https://doi.org/10.1016/j.eap.2024.08.010

[9] Huang, Z. (2026). Crowding-out effect of digitalized tax administration on corporate green innovation: Evidence from China's Golden Tax Phase IV. *Finance Research Letters*, *87*, 109067. https://doi.org/10.1016/j.frl.2025.109067

[10] State Taxation Administration of the People's Republic of China. (2024). *Announcement on the promotion and application of fully digitalized electronic invoices* (Announcement No. 11 of 2024). https://fgk.chinatax.gov.cn/zcfgk/c100012/c5236067/content.html

[11] Kothari, S. P., Leone, A. J., & Wasley, C. E. (2005). Performance matched discretionary accrual

measures. *Journal of Accounting and Economics*, *39*(1), 163–197. https://doi.org/10.1016/j.jacceco.2004.11.002

[12] Roychowdhury, S. (2006). Earnings management through real activities manipulation. *Journal of Accounting and Economics*, *42*(3), 335–370. https://doi.org/10.1016/j.jacceco.2006.01.002

[13] Cohen, D. A., & Zarowin, P. (2010). Accrual-based and real earnings management activities around seasoned equity offerings. *Journal of Accounting and Economics*, *50*(1), 2–19. https://doi.org/10.1016/j.jacceco.2010.01.002

[14] Shenzhen CSMAR Data Technology. (2025). *China Stock Market and Accounting Research database* [Data set]. https://data.csmar.com

[15] Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, *225*(2), 254–277. https://doi.org/10.1016/j.jeconom.2021.03.014

[16] Callaway, B., & Sant'Anna, P. H. C. (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics*, *225*(2), 200–230. https://doi.org/10.1016/j.jeconom.2020.12.001

[17] Sun, L., & Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, *225*(2), 175–199. https://doi.org/10.1016/j.jeconom.2020.09.006

[18] Borusyak, K., Jaravel, X., & Spiess, J. (2024). Revisiting event-study designs: Robust and efficient estimation. *The Review of Economic Studies*, *91*(6), 3253–3285. https://doi.org/10.1093/restud/rdae007

[19] Cameron, A. C., Gelbach, J. B., & Miller, D. L. (2008). Bootstrap-based improvements for inference with clustered errors. *The Review of Economics and Statistics*, *90*(3), 414–427. https://doi.org/10.1162/rest.90.3.414

[20] Roth, J. (2022). Pretest with caution: Event-study estimates after testing for parallel trends. *American Economic Review: Insights*, *4*(3), 305–322. https://doi.org/10.1257/aeri.20210236

WISDOM ACADEMIC https://doi.org/10.63808/ftd.v2i3.487 Volume 2 , Issue 3

Appendix A. Variable definitions and pilot launch dates

Table A1

Definition and source of each variable

Variable	Definition	Source
|DA|	Absolute residual from equation (1), by industry and year	CSMAR
REM	Composite measure from equation (2)	CSMAR
GTPIV	One from the year the firm's jurisdiction launches the issuing pilot	STA
Size	Natural logarithm of total assets	CSMAR
Lev	Total liabilities over total assets	CSMAR
ROA	Net income over total assets	CSMAR
Growth	Annual growth rate of operating revenue	CSMAR
CFO	Operating cash flow over total assets	CSMAR
Top1	Shareholding of the largest shareholder	CSMAR
Indep	Independent directors over board size	CSMAR
Dual	One if the chief executive chairs the board	CSMAR
Big4	One if audited by an international Big Four firm	CSMAR
MB	Market value over book value of equity	CSMAR
Loss	One if net income is negative	CSMAR
SOE	One if the ultimate controller is a government agency or state-owned entity	CSMAR
HighAudit	One if audited by an international Big Four firm or a CICPA top-ten domestic auditor	CSMAR; CICPA
InpInt	Two-digit industry median ratio of operating costs to operating revenue, split at the sample median	CSMAR
ANAFERR	Absolute median forecast error over share price	CSMAR
FDISP	Standard deviation of analyst forecasts over share price	CSMAR

BTD	Pre-tax accounting income less estimated taxable income, over total assets	CSMAR
DDBTD	Residual from regressing BTD on total accruals [2]	CSMAR

Note. Estimated taxable income is income tax expense less deferred tax expense over the applicable statutory rate. CSMAR is the China Stock Market and Accounting Research database [14], STA the State Taxation Administration and CICPA the Chinese Institute of Certified Public Accountants.

Table A2

Launch of the fully digitalized e-invoice issuing pilot by provincial-level tax jurisdiction

Launch date of issuing pilot	Provincial-level tax jurisdictions	Number	Assigned onset cohort
1 Dec 2021	Guangdong (excluding Shenzhen), Shanghai, Inner Mongolia	3	2022
28 Oct 2022	Sichuan	1	2023
28 Nov 2022	Xiamen	1	2023
Jan-Mar 2023	Tianjin, Chongqing, Shaanxi, Dalian, Ningbo, Qingdao, Shenzhen, Yunnan, Fujian, Henan, Jilin	11	2023
27 Apr 2023	Shanxi, Liaoning, Jiangsu, Zhejiang, Jiangxi, Hainan, Gansu, Guangxi	8	2023
1 Oct 2023	Hebei, Hubei, Heilongjiang, Xinjiang	4	2024
1 Nov 2023	Beijing, Anhui, Shandong, Hunan, Guizhou, Qinghai, Ningxia	7	2024
Dec 2023	Tibet	1	2024

Note. Dates refer to the issuing pilot. The separate receiving pilot reached all thirty-six jurisdictions on 28 August 2022 and is not used to date treatment. The eleven jurisdictions launching between January and March 2023 share a cohort under the half-year rule. Nationwide extension from 1 December 2024 [10] lies outside every onset window, and announcement numbers by jurisdiction are in the replication file.