

Disclosure Reform, Auditable LLM Measurement, and Post-Disclosure Trading Response: Evidence from Chinese Listed Securities Companies

Haoyu Wang

*School of Economics and Management, Beijing Jiaotong University, Beijing, China
23725065@bjtu.edu.cn*

Abstract. This paper studies whether the 2021 CSRC disclosure reform package changed forward-looking risk disclosure and investor trading of Chinese listed securities companies. It combines an auditable LLM measurement framework with a continuous difference-in-differences (DID) design. Treatment intensity is fixed from 2015-2020 annual-report text and captures pre-policy boilerplate dependence through the standardized `insufficient_evidence` share and `low-RiskForward` share in section-level LLM records. In 307 firm-years for the disclosure model and 306 firm-years for the market-response model, high-exposure brokers raise `RiskForward` more after the reform and record higher `TurnoverPost` from the announcement date to year-end. A one-standard-deviation increase in treatment intensity maps to a 0.570-point rise in `RiskForward` and a 0.965-point rise in `TurnoverPost`, or 18.1% and 14.7% of regression-sample means. Dynamic DID estimates show no detectable differential pre-trend before 2021, and pre-period placebo tests remain null. Channel diagnostics are consistent with a disclosure mechanism, but they do not establish one uniquely. The evidence thus supports a policy-consistent reading that reform-related improvement in forward-looking risk disclosure is associated with stronger investor trading, subject to standard DID assumptions.

Keywords: Large Language Models, Disclosure Reform, Forward-Looking Risk Disclosure, Stock Turnover, Difference-in-Differences

1. Introduction

Chinese listed securities companies are important capital-market intermediaries, and their annual reports provide outside investors with structured information about market risk, liquidity pressure, capital adequacy, compliance exposure, and business sustainability.

The 2021 CSRC disclosure reform package makes this setting especially relevant. The revised Measures for the Administration of Information Disclosure by Listed Companies and the revised annual-report content and format standards strengthened periodic-report requirements [1, 2], creating a direct institutional background for examining whether broker risk discussion became less boilerplate and more decision-useful.

This paper asks whether the 2021 CSRC disclosure reform changed forward-looking risk disclosure by Chinese listed securities companies and whether brokers that were more exposed to pre-policy boilerplate disclosure showed stronger post-disclosure investor trading responses after the reform.

The question sits at the intersection of disclosure regulation, financial text measurement, and securities-market response. Prior disclosure studies show that reporting quality, readability, tone, and risk wording matter for market outcomes [3-12]. However, standard textual proxies do not isolate substantive forward-looking risk disclosure very well, and contemporaneous firm-year regressions remain vulnerable to simultaneity and omitted-variable concerns. Recent LLM studies broaden what can be measured semantically, but they often emphasize model capability rather than an auditable measurement process embedded in a policy-based design [13-17].

The empirical design combines an evidence-constrained LLM scoring workflow with a continuous difference-in-differences design. Treatment intensity is defined before the policy shock: BoilerplateDependence is constructed from 2015-2020 annual-report text and captures reliance on low-information risk language through insufficient_evidence records and low RiskForward scores. The market outcome, TurnoverPost, is measured only after the annual report becomes public.

The sample contains 43 Chinese A-share listed securities companies over report years 2015-2023. This single-industry setting is useful because listed brokers face a common regulatory environment, have relatively comparable disclosure obligations, and provide public trading data that allow the paper to connect annual-report disclosure with post-disclosure investor activity.

The paper contributes in three ways. First, it studies disclosure reform in a regulated financial-intermediary industry where risk discussion is central to investor assessment. Second, it uses an auditable LLM framework to measure forward-looking risk disclosure while retaining unsupported cases as observable low-information records. Third, it links policy-induced disclosure variation to post-disclosure trading response without treating the LLM itself as a causal engine.

The remainder of the paper is organized as follows. Section 2 reviews the literature and develops the hypothesis. Section 3 describes the policy shock, scoring framework, sample, variables, and empirical design. Section 4 reports the empirical results. Section 5 concludes.

2. Literature review and hypothesis development

2.1. Disclosure regulation, textual reporting, and market behavior

Classic disclosure research argues that better disclosure improves the information environment and changes capital-market outcomes, while later work shows that readability, tone, and risk wording shape how investors process filings [3-7, 12]. These studies establish that disclosure text can influence market behavior, but they rarely use an external disclosure-policy shock to test whether such effects intensify in a regulated financial-intermediary industry.

2.2. Risk disclosure, boilerplate language, and forward-looking content

Risk disclosure is economically important because it communicates uncertainty about future firm conditions. This literature distinguishes boilerplate risk mention from genuinely informative forward-looking disclosure [8-11]. Tao et al. show that forward-looking language can be identified as a distinct informational object, but prior work does not provide an auditable approach to forward-looking broker risk disclosure in China or link that measurement to a regulatory shock [18].

2.3. LLM-based financial text measurement and auditability

Recent LLM research in finance shows that large models can classify, summarize, and evaluate financial text at a richer semantic level than traditional dictionaries [13-17]. Empirical disclosure research also needs auditability. This paper therefore uses the LLM as a measurement instrument with fixed prompts, a fixed output schema, retained evidence, and explicit unsupported cases, rather than as a free-form explanation engine.

2.4. Hypothesis development

The 2021 CSRC disclosure reform package tightened listed-company disclosure governance and revised the annual-report format standards [1, 2]. The reform increased regulatory pressure for periodic reports to provide targeted and decision-useful narrative information rather than purely formal compliance text. This pressure should matter most for firms whose pre-policy risk sections relied heavily on low-information, weakly supported, or boilerplate-like language because those firms had greater room to improve the substance of their disclosures.

For listed securities companies, the predicted response is also economically meaningful. Broker risk disclosure concerns market-risk exposure, capital pressure, compliance events, business concentration, and risk-management responses. If the reform induces high-boilerplate brokers to provide more specific forward-looking risk discussion, investors receive more decision-relevant information after the annual report is released and may trade more actively while processing that information. The hypothesis therefore links the disclosure-stage response to a post-disclosure market-response prediction, while the empirical tests interpret the channel evidence cautiously rather than as proof of a unique mediation path.

Accordingly, the paper tests the following hypothesis:

H1: After the 2021 disclosure reform, listed securities companies with higher pre-policy boilerplate dependence experience larger increases in forward-looking risk disclosure and stronger post-disclosure investor trading response.

3. Research design

3.1. Policy shock and identifying variation

The identification strategy uses the 2021 CSRC disclosure reform package as a common external shock. Two official documents are central. First, the revised *Measures for the Administration of Information Disclosure by Listed Companies* were promulgated by the CSRC in March 2021 and took effect on May 1, 2021 [1]. Second, the revised annual-report content and format standards were issued in June 2021 [2]. Together, these measures tightened the disclosure framework governing listed-company periodic reports. The reform is particularly relevant to listed brokers because annual reports are one of the main formal channels through which they disclose changes in market risk, capital pressure, business concentration, compliance conditions, and risk management under a strongly supervised regime. Because the annual-report format revision was issued only in mid-2021, the first fully treated reporting cycle in the present design is the fiscal-year 2021 annual report, which was disclosed in 2022. Accordingly, the paper defines Post = 1 for report years 2021 and later.

The treatment is not binary at the firm level. Instead, the design uses a predetermined continuous exposure measure, BoilerplateDependence, constructed only from 2015-2020 text. This avoids defining treatment with post-policy disclosure outcomes. Firms with higher BoilerplateDependence

are those whose pre-policy annual-report sections more often contained low-information or weakly supported forward-looking risk discussion. The identifying assumption is therefore that, absent the reform, high- and low-exposure brokers would not have shown systematically different post-2021 changes in RiskForward and post-disclosure turnover.

3.2. Auditable LLM scoring framework

Annual-report PDFs are collected from CNINFO, converted to text, and segmented into high-information passages, mainly management discussion and analysis, major-events sections, and risk-related sections. The preprocessing workflow excludes low-information template material such as cover-page statements, tables of contents, definitions pages, board responsibility statements, audit-report text, and generic legal disclaimers. Within retained narrative sections, generic risk lists that merely repeat standard categories without firm-specific future exposure, likely impact, or response measures are not treated as strong forward-looking risk disclosure.

Each retained section is scored by an evidence-constrained LLM prompt under a fixed flat JSON schema. The archived empirical run used DeepSeek Chat (deepseek-chat) with temperature = 0.2 and JSON-object response formatting. Because that run did not document a fixed random seed, the paper does not claim bit-for-bit deterministic regeneration; auditability rests on prompt and schema disclosure, archived raw JSON outputs, parsed scores, and validation checks.

The focal score is RiskForward, the firm-year average of section-level forward-looking risk scores on a 1-5 scale. Unsupported cases are not filled with speculative text. They are flagged as `insufficient_evidence` and retained as low-information observations; in the archived scoring data, those records receive the minimum section-level forward-looking risk score of 1. This rule is important for the policy design because it allows the pre-policy treatment intensity to reflect low-information disclosure rather than silently dropping it.

3.3. Sample and variables

The sample is constructed from Chinese A-share listed securities companies. The initial text-based annual-report panel contains 361 firm-years for 43 listed brokers over report years 2015-2024. Because the policy design requires post-disclosure market outcomes that can be observed after the annual report becomes public, the analysis restricts the main report-year window to 2015-2023. Matching annual-report announcement dates and post-disclosure turnover produces 318 policy-panel firm-years. The first-stage disclosure regression uses 307 observations after requiring complete controls and pre-policy exposure, and the reduced-form turnover regression uses 306 observations after also requiring the post-disclosure turnover outcome.

The listed-securities-company sample is appropriate for three reasons. First, brokers operate under a strong and relatively common regulatory environment, which reduces cross-industry heterogeneity in disclosure incentives. Second, the 2021 CSRC disclosure reform is directly relevant to their annual reports because these filings discuss market risk, capital pressure, business concentration, compliance conditions, and risk management. Third, listed brokers provide both public annual-report text and market trading data, making it possible to study disclosure change and investor trading response in the same firm-year panel.

The key variables are defined as follows. RiskForward is the firm-year average of section-level forward-looking risk scores. BoilerplateDependence is a firm-level pre-policy exposure measure, defined as the average of two standardized 2015-2020 components: the share of section-level records flagged as `insufficient_evidence` and the share of section-level records receiving low

forward-looking risk scores. TurnoverPost is average weekly turnover from the annual-report announcement date to the end of the disclosure year. SigmaPost is the standard deviation of weekly returns over the same post-disclosure interval. Controls include Size, Lev, BM, ROA, Sigma, Ret, and Age, measured at the firm-year level and winsorized at the 1st and 99th percentiles.

The use of TurnoverPost rather than same-calendar-year average turnover is deliberate. A same-year turnover measure would mix weeks before the annual report becomes public with weeks after disclosure, which weakens the temporal ordering required for a reviewer-friendly causal interpretation. By averaging turnover only from the announcement date to year-end, the outcome window starts after the filing is released and therefore aligns the dependent variable more closely with the information arrival generated by the annual report. This construction does not eliminate all identification concerns, but it is cleaner than the earlier year-average turnover measure for the present policy design.

3.4. Empirical models

The first-stage disclosure model is:

$$\begin{aligned} RiskForward_{i,t} = & \alpha + \beta(BoilerplateDependence_i \times Post_t) \\ & + \gamma Controls_{i,t} + FirmFE_i + YearFE_t + \varepsilon_{i,t} \end{aligned} \quad (1)$$

The reduced-form market-response model is:

$$\begin{aligned} TurnoverPost_{i,t} = & \alpha + \beta(BoilerplateDependence_i \times Post_t) \\ & + \gamma Controls_{i,t} + FirmFE_i + YearFE_t + \varepsilon_{i,t} \end{aligned} \quad (2)$$

Firm fixed effects absorb time-invariant differences across brokers, and year fixed effects absorb common market and regulatory shocks. Standard errors are clustered at the firm level following standard difference-in-differences practice [19]. To examine whether identification is plausible, the paper also reports dynamic DID coefficients and pre-period placebo tests. To assess whether the result is concentrated where formal disclosure matters more, the paper reports split-sample evidence for smaller and larger brokers based on pre-policy average size.

4. Empirical results

4.1. Descriptive evidence

Table 1 reports descriptive statistics for the policy-DID sample. Mean TurnoverPost is 7.815, mean RiskForward is 3.147, and BoilerplateDependence varies meaningfully across firms. Its two components, PreInsuffShare and PreLowRFShare, capture persistent cross-firm differences in low-information disclosure before the reform.

Table 1. Descriptive statistics

Var.	N	Mean	SD	Min	P25	Median	P75	Max	Miss.
TurnoverPost	318	7.815	10.117	0.431	3.170	4.745	7.618	82.231	0
RiskForward	318	3.147	1.031	1.000	2.333	3.000	4.000	5.000	0
BoilerplateDependence	311	0.049	0.745	-0.395	-0.395	-0.395	0.287	3.149	7
PreInsuffShare	311	0.005	0.028	0.000	0.000	0.000	0.000	0.167	7
PreLowRFShare	311	0.172	0.253	0.000	0.000	0.000	0.167	1.000	7
Size	318	25.586	1.002	23.620	24.806	25.314	26.392	27.916	0
Lev	318	0.743	0.062	0.495	0.720	0.752	0.783	0.844	0
BM	312	0.074	0.034	0.015	0.049	0.073	0.096	0.178	6
ROA	318	0.016	0.008	-0.003	0.011	0.015	0.019	0.039	0

Note: This table reports summary statistics for the policy-DID sample used in the disclosure-reform analysis.

The sample-selection path starts from the 361 firm-year text panel, retains report years through 2023 so that post-disclosure market outcomes are observable, and then keeps firm-years with matched annual-report announcement dates and post-disclosure turnover. The final regression samples contain 307 and 306 observations, respectively, for the disclosure and turnover equations.

4.2. First-stage effect on forward-looking risk disclosure

Table 2 first reports the disclosure-stage DID result. The coefficient on BoilerplateDependence x Post is 0.762 with a clustered p-value of 0.023. Thus, after the 2021 reform package, firms with higher pre-policy boilerplate dependence experienced larger increases in RiskForward, as predicted by the disclosure-reform design.

Table 2. Main DID estimates

Model	N	R2	Coef.	p
First stage: RiskForward	307	0.590	0.762**	0.0232
Reduced form: TurnoverPost	306	0.694	1.289**	0.0354
Supplementary: SigmaPost	306	0.569	0.003*	0.0912

Note: *, **, and *** indicate $p < 0.10$, $p < 0.05$, and $p < 0.01$.

Note: This table reports the first-stage disclosure DID, the reduced-form turnover DID, and the supplementary post-disclosure volatility specification.

4.3. Reduced-form effect on investor trading response

Table 2 next reports the reduced-form market-response model. The coefficient on BoilerplateDependence x Post in the TurnoverPost regression is 1.289 with a p-value of 0.035. High-exposure firms therefore experienced stronger trading after annual-report disclosure in the post-reform period.

Table 3 shows that the market-response result is not driven by one coding choice. The turnover coefficient remains positive when using raw PreInsuffShare (0.337, $p = 0.001$), the alternative low-RiskForward exposure (1.313, $p = 0.043$), or a 2022 post-policy cutoff (1.614, $p = 0.044$).

Table 3. Robustness and placebo tests

Model	N	R2	Coef.	p
First stage: PreInsuffShare	307	0.531	0.148***	0.0003
Turnover: PreInsuffShare	306	0.690	0.337***	0.0011
First stage: LowRiskForwardExposure	307	0.649	0.779***	0.0000
Turnover: LowRiskForwardExposure	306	0.699	1.313**	0.0426
First stage: Post2022	307	0.555	0.580**	0.0211
Turnover: Post2022	306	0.696	1.614**	0.0440
Placebo 2018: RiskForward	187	0.716	0.156	0.2743
Placebo 2018: TurnoverPost	187	0.808	-0.423	0.4466
Placebo 2019: RiskForward	187	0.717	0.183	0.2842
Placebo 2019: TurnoverPost	187	0.808	0.333	0.6802

Note: *, **, and *** indicate $p < 0.10$, $p < 0.05$, and $p < 0.01$.

Note: This table reports alternative exposure definitions, an alternative post-period cutoff, and pre-period placebo tests.

The magnitude is also economically interpretable. Table 4 shows that a one-standard-deviation increase in BoilerplateDependence is associated with a 0.570-point increase in RiskForward, equal to 18.1 percent of the RiskForward sample mean and 14.2 percent of the 1-5 scoring range. For TurnoverPost, the same change corresponds to a 0.965-point increase, or 14.7 percent of the regression-sample mean.

Table 4. Economic significance

Outcome	Coef.	Exposure SD	1 SD Effect	Mean	% Mean	% Score Range
RiskForward	0.762	0.748	0.570	3.156	18.052	14.245
TurnoverPost	1.289	0.749	0.965	6.586	14.656	

Note: This table reports the implied effect of a one-standard-deviation increase in pre-policy boilerplate dependence.

As a supplementary outcome, the SigmaPost coefficient is positive but weaker (0.0025, $p = 0.091$), suggesting that the main evidence is strongest for trading intensity.

4.4. Disclosure-channel diagnostics

Table 5 reports a direct set of disclosure-channel diagnostics.

The first-stage result shows that policy exposure increases RiskForward (0.762, $p = 0.023$), and the reduced-form result shows that the same exposure increases TurnoverPost (1.289, $p = 0.035$). Together, these estimates provide the main empirical support for the disclosure-channel interpretation.

Table 5. Disclosure-channel diagnostics

Model	Term	N	R2	Coef.	SE	p
First stage: policy exposure to RiskForward	BoilerPost	307	0.590	0.762**	0.322	0.0232

Table 5. (continued)

Reduced form: policy exposure to TurnoverPost	BoilerPost	306	0.694	1.289**	0.591	0.0354
Disclosure score in turnover equation	RiskForward	311	0.724	0.563	0.353	0.1185
Policy exposure after adding RiskForward	BoilerPost	306	0.696	1.015*	0.577	0.0863
RiskForward after adding policy exposure	RiskForward	306	0.696	0.359	0.379	0.3490
Short-window trading response	BoilerPost	306	0.695	1.612*	0.877	0.0736
Post-disclosure volatility response	BoilerPost	306	0.569	0.003*	0.001	0.0912

Note: *, **, and *** indicate $p < 0.10$, $p < 0.05$, and $p < 0.01$.

Note: This table reports first-stage, reduced-form, mediator-diagnostic, short-window turnover, and post-disclosure volatility specifications. The mediator rows are interpreted as channel diagnostics rather than definitive causal mediation estimates.

The table also examines whether the measured disclosure score is associated with trading. Replacing policy exposure with RiskForward yields a positive but insignificant coefficient (0.563, $p = 0.118$). Including both BoilerplateDependence $\times$ Post and RiskForward reduces the policy coefficient from 1.289 to 1.015, while RiskForward remains imprecisely estimated. The pattern is consistent with partial disclosure-channel evidence, but it is not a definitive mediation test.

The short-window and volatility diagnostics point in the same direction: the 12-week turnover coefficient is 1.612 ($p = 0.074$), and the volatility response is 0.0025 ($p = 0.091$). Overall, the channel evidence supports the disclosure interpretation, while the paper remains cautious about mechanism strength.

4.5. Dynamic DID and placebo evidence

Figure 1 plots the dynamic DID coefficients for RiskForward. The pre-reform coefficients for 2016-2019 are statistically insignificant, whereas the 2021 and 2022 coefficients are positive and significant. The pattern is more consistent with reform-related divergence than with a persistent pre-existing gap.

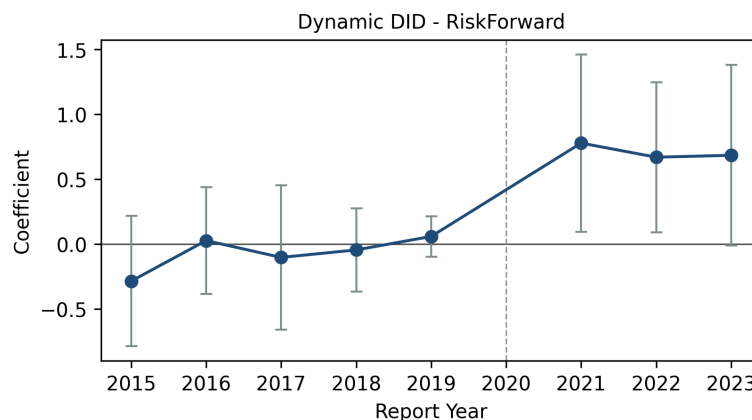


Figure 1. Dynamic DID - RiskForward

Note: This figure plots year-specific coefficients for BoilerplateDependence interacted with report-year dummies, with 2020 omitted as the base year.

Figure 2 re-plots the dynamic DID coefficients for TurnoverPost as point estimates with 95 percent confidence intervals. The pre-reform coefficients are statistically indistinguishable from zero, although the turnover dynamics are visually noisier than the disclosure-stage dynamics in Figure 1. The post-reform estimates turn positive and are larger in 2022-2023, with the 2023 estimate marginally significant. The figure is therefore interpreted as a diagnostic check, not as mechanical proof of the identifying assumption.

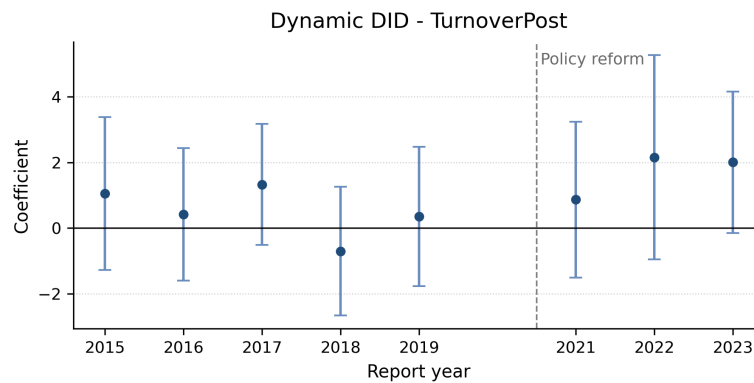


Figure 2. Dynamic DID - turnoverpost

Note: This figure plots year-specific coefficients for BoilerplateDependence interacted with report-year dummies in the post-disclosure turnover regression, with 2020 omitted as the base year.

Table 3 also reports placebo tests using only the pre-2021 sample and imposing fake reform dates in 2018 and 2019. The RiskForward and TurnoverPost placebo coefficients are statistically insignificant, which reduces concern that the main estimates are driven by strong differential pre-trends while still leaving the identifying assumption as a maintained research-design condition.

4.6. Formal interaction heterogeneity tests

Table 6 replaces a purely split-sample discussion with formal interaction specifications. The turnover-stage interaction for smaller brokers is not statistically different from zero, and the interaction terms for low institutional ownership and high prior violation exposure are also imprecise. The split-sample patterns should therefore be read as directional rather than as sharply identified cross-group differences.

Table 6. Formal interaction heterogeneity tests

Model	Outcome	N	R2	Base Term	Base Coef.	Base SE	Base p	Interaction	Interaction Coef.	Interaction SE	Interaction p
Disclosure stage x Smaller brokers	RiskForward	307	0.609	BoilerPost	1.407***	0.389	0.001	BoilerPost: SmallBroker	-0.909*	0.451	0.051
Turnover stage x Smaller brokers	TurnoverPost	306	0.694	BoilerPost	1.241	0.995	0.220	BoilerPost: SmallBroker	0.068	1.323	0.959

Table 6. (continued)

Turnover stage x Low institutional ownership	TurnoverPost	306	0.699	BoilerPost	4.057	2.701	0.141	BoilerPost: LowPreInst	-3.255	2.766	0.246
Turnover stage x High prior violation exposure	TurnoverPost	306	0.698	BoilerPost	4.576	3.994	0.259	BoilerPost: HighPreViolation	-3.684	4.115	0.376

Note: *, **, and *** indicate $p < 0.10$, $p < 0.05$, and $p < 0.01$.

Note: This table reports formal interaction models for broker size, low institutional ownership, and high prior violation exposure. The interaction terms test whether the reform effect differs significantly across groups.

At the same time, the first-stage disclosure equation shows clearer heterogeneity: the baseline reform effect remains positive and strongly significant, while the interaction with SmallBroker is negative and marginally significant. Heterogeneity is therefore clearer at the disclosure-adjustment stage than at the trading stage.

4.7. Split-sample and supportive environment evidence

Table 7 reports the original size-split reduced-form evidence, and Figure 3 visualizes the same pattern. The coefficient on BoilerplateDependence x Post is positive and significant among smaller brokers (1.627, $p = 0.037$), but positive and insignificant among larger brokers (1.223, $p = 0.246$). This pattern is economically plausible in the Chinese broker setting because smaller brokers usually have fewer alternative information channels and weaker perceived implicit support, so formal annual reports carry more weight when disclosure standards tighten. Even so, Table 6 indicates that the cross-group contrast should be read as suggestive rather than definitive.

Table 7. Split-sample evidence by broker size

Group	N	Coef.	SE	p
Smaller brokers	142	1.627**	0.725	0.0368
Larger brokers	164	1.223	1.021	0.2457

Note: *, **, and *** indicate $p < 0.10$, $p < 0.05$, and $p < 0.01$.

Note: This table reports reduced-form DID coefficients separately for smaller and larger brokers based on pre-policy average size.

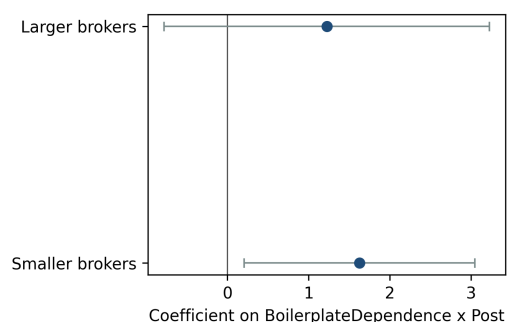


Figure 3. Reduced-form DID by broker size

Note: This figure plots the reduced-form DID coefficient for smaller and larger brokers separately.

Tables 8 and 9 summarize supportive environment evidence. The main reduced-form coefficient stays positive and significant after adding institutional ownership and violation controls, and split-sample extensions show stronger responses in low-ownership and high-violation environments. Because the formal interaction terms for these dimensions are not sharply estimated, the paper treats these results as contextual support rather than stand-alone mechanism proof.

Table 8. Additional controls: institutional ownership and violations

N	R2	Coef.	p	Model
306	0.695	1.246**	0.0402	Turnover + institutional ownership control
306	0.695	1.273**	0.0374	Turnover + violation control
306	0.695	1.233**	0.0429	Turnover + both controls
307	0.595	0.762**	0.0208	RiskForward + both controls

Note: *, **, and *** indicate $p < 0.10$, $p < 0.05$, and $p < 0.01$.

Note: This table reports additional DID specifications that control for institutional ownership and listed-company violation events.

Table 9. Supportive environment split-sample extensions

Split	Group	N	Coef.	p
Low institutional ownership	Low institutional ownership: Yes	170	0.943**	0.0185
Low institutional ownership	Low institutional ownership: No	136	5.274	0.1002
High prior violation exposure	High prior violation exposure: Yes	194	0.777**	0.0394
High prior violation exposure	High prior violation exposure: No	112	6.236	0.1318

Note: *, **, and *** indicate $p < 0.10$, $p < 0.05$, and $p < 0.01$.

Note: This table reports reduced-form DID coefficients separately by pre-policy institutional ownership and pre-policy violation exposure.

4.8. Expanded robustness checks

Table 10 broadens the robustness analysis beyond the earlier placebo and alternative-exposure checks. The baseline reduced-form coefficient remains significant after outlier treatment, alternative timing cutoffs, excluding selected years, and replacing the treatment with either raw PreInsuffShare or low-RiskForward exposure. A shorter 12-week post-disclosure outcome remains positive, although only marginally significant. Overall, the result is not tied to a single exposure definition, turnover construction, or small set of extreme observations.

Table 10. Expanded robustness summary

Model	Term	N	R2	Coef.	SE	p
Baseline replication	BoilerPost	306	0.694	1.289**	0.591	0.0354
Winsorized turnover outcome	BoilerPost	306	0.694	1.286**	0.590	0.0352
12-week post-disclosure turnover	BoilerPost	306	0.695	1.612*	0.877	0.0736
Trim top/bottom 1% turnover	BoilerPost	302	0.694	1.338**	0.595	0.0302
Exclude report year 2020	BoilerPost	266	0.740	0.808**	0.349	0.0260

Table 10. (continued)

Exclude first treated cycle 2021	BoilerPost	266	0.744	1.660**	0.743	0.0313
Later policy cutoff (Post2022)	BoilerPost2022	306	0.696	1.614**	0.776	0.0440
Alternative exposure: PreInsuffShare	InsuffPost	306	0.690	0.337***	0.096	0.0011
Alternative exposure: LowRiskForwardExposure	LowRFPost	306	0.699	1.313**	0.626	0.0426
Add institutional ownership and violations	BoilerPost	306	0.695	1.233**	0.593	0.0441

Note: *, **, and *** indicate $p < 0.10$, $p < 0.05$, and $p < 0.01$.

Note: This table reports robustness checks using alternative outcome constructions, outlier treatment, alternative timing cutoffs, alternative exposure definitions, and richer control sets.

Table 11 reports standard-error robustness checks for the main reduced-form turnover model. The coefficient on BoilerplateDependence x Post remains positive and statistically significant under heteroskedasticity-robust HC1 standard errors, firm-clustered standard errors, year-clustered standard errors, and two-way firm-year clustered

standard errors. Because the sample contains a limited number of report years, the year-clustered and two-way clustered estimates are best viewed as sensitivity checks rather than replacements for the preferred firm-clustered specification. Nevertheless, the statistical inference is not dependent on a single standard-error convention.

Table 11. Standard-error robustness checks

Model	Term	N	R2	Coef.	SE	p
Heteroskedasticity-robust HC1	BoilerPost	306	0.694	1.289**	0.513	0.0127
Firm-clustered SE	BoilerPost	306	0.694	1.289**	0.588	0.0343
Year-clustered SE	BoilerPost	306	0.694	1.289**	0.462	0.0236
Two-way firm-year clustered SE	BoilerPost	306	0.694	1.289**	0.544	0.0452

Note: *, **, and *** indicate $p < 0.10$, $p < 0.05$, and $p < 0.01$.

Note: This table reports the main reduced-form turnover coefficient under alternative standard-error conventions. Year and two-way clustered estimates are sensitivity checks because the panel has a limited number of years.

4.9. Discussion and identification limits

Overall, the evidence is consistent with the intended disclosure channel. Firms with higher pre-policy boilerplate dependence changed their forward-looking risk disclosure more strongly after the reform and then exhibited stronger post-disclosure trading. The design improves on a simple year-frequency association because treatment intensity is predetermined and the market outcome is measured after the annual report is disclosed.

At the same time, the estimates should be read as policy-consistent DID evidence rather than definitive structural proof. The design still depends on a parallel-trends assumption and on a policy package rather than a single micro-level regulatory action, so the paper emphasizes dynamic diagnostics, placebo tests, robustness checks, and cautious language around heterogeneity.

5. Conclusion

The design uses a documented policy shock, defines treatment intensity from pre-policy text, and measures market response after the annual report is disclosed. This paper studies whether a disclosure reform shock changed forward-looking risk disclosure and investor trading in Chinese listed securities companies.

The same firms also have higher post-disclosure turnover. The results show that the 2021 CSRC disclosure reform package increased RiskForward more strongly for brokers with higher pre-policy boilerplate dependence. The dynamic DID estimates show no detectable pre-reform differential trend, and the pre-period placebo tests are null. A one-standard-deviation increase in treatment intensity is associated with a 0.570-point increase in RiskForward and a 0.965-point increase in TurnoverPost. The turnover result survives alternative exposure definitions, timing cutoffs, outlier treatment, richer controls, and alternative standard-error conventions. The heterogeneity patterns are suggestive, and the channel diagnostics support but do not fully establish the disclosure mechanism.

These findings show how an auditable LLM framework can be embedded in a policy-based empirical design rather than used only for descriptive text scoring. They also suggest that disclosure regulation can alter the informational role of annual-report risk discussion in a regulated financial-intermediary industry.

It would be possible to investigate in future research whether the disclosure reform changes brokers' later risk-taking behavior and whether the finer event-window outcomes make the market response sharper.

References

- [1] China Securities Regulatory Commission. (2021). Measures for the administration of information disclosure by listed companies (CSRC Order No. 182) [in Chinese]. *China Securities Regulatory Commission*. <https://www.csrc.gov.cn/csrc/c101864/c2ee1a791fddc4f5ebee70aa8e2399cf/content.shtml>
- [2] China Securities Regulatory Commission. (2021). Standards for the contents and formats of information disclosure by companies offering securities to the public No. 2: Contents and formats of annual reports (2021 revision) [in Chinese]. *China Securities Regulatory Commission*. <https://www.csrc.gov.cn/csrc/c101864/c6df1268b5b294448bdec7e010d880a01/content.shtml>
- [3] Botosan, C. A. (1997). Disclosure level and the cost of equity capital. *The Accounting Review*, 72(3), 323-349. <https://cir.nii.ac.jp/crid/1370306905175703580>
- [4] Leuz, C., & Verrecchia, R. E. (2000). The economic consequences of increased disclosure. *Journal of Accounting Research*, 38(Supplement 1), 91-124. <https://ideas.repec.org/a/bla/joares/v38y2000isupplementp91-124.html>
- [5] Li, F. (2008). Annual report readability, current earnings, and earnings persistence. *Journal of Accounting and Economics*, 45(2-3), 221-247. doi:10.1016/j.jacceco.2008.02.003. <https://doi.org/10.1016/j.jacceco.2008.02.003>
- [6] Loughran, T., & McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *The Journal of Finance*, 66(1), 35-65. <https://doi.org/10.1111/j.1540-6261.2010.01625.x>
- [7] Loughran, T., & McDonald, B. (2016). Textual analysis in accounting and finance: A survey. *Journal of Accounting Research*, 54(4), 1187-1230. <https://doi.org/10.1111/1475-679X.12123>
- [8] Kravet, T., & Muslu, V. (2013). Textual risk disclosures and investors' risk perceptions. *Review of Accounting Studies*, 18(4), 1088-1122. <https://doi.org/10.1007/s11142-013-9228-9>
- [9] Campbell, J. L., Chen, H., Dhaliwal, D. S., Lu, H. M., & Steele, L. B. (2014). The information content of mandatory risk factor disclosures in corporate filings. *Review of Accounting Studies*, 19(1), 396-455. <https://doi.org/10.1007/s11142-013-9258-3>
- [10] Elshandidy, T., Shrivs, P. J., Bamber, M., & Abraham, S. (2018). Risk reporting: A review of the literature and implications for future research. *Journal of Accounting Literature*, 40, 54-82. <https://doi.org/10.1016/j.acclit.2017.12.001>
- [11] Khlif, H., & Hussainey, K. (2016). The association between risk disclosure and firm characteristics: A meta-analysis. *Journal of Risk Research*, 19(2), 181-211. <https://doi.org/10.1080/13669877.2014.961514>

- [12] Dang, M., Puwanenthiren, P., Nguyen, M. T., Hoang, V. A., Mazur, M., & Henry, D. (2022). Does managerial tone matter for stock liquidity? Evidence from textual disclosures. *Finance Research Letters*, 48, 102917. <https://doi.org/10.1016/j.frl.2022.102917>
- [13] Wu, S., Irsoy, O., Lu, S., Dabrovolski, V., Dredze, M., Gehrmann, S., Kambadur, P., Rosenberg, D., & Mann, G. (2023). BloombergGPT: A large language model for finance. *arXiv preprint arXiv:2303.17564*. <https://arxiv.org/abs/2303.17564>
- [14] Yang, H., Liu, X. Y., & Wang, C. D. (2023). FinGPT: Open-source financial large language models. *arXiv preprint arXiv:2306.06031*. <https://arxiv.org/abs/2306.06031>
- [15] Xie, Q., Han, W., Zhang, X., Lai, Y., Peng, M., Lopez-Lira, A., & Huang, J. (2023). PIXIU: A comprehensive benchmark, instruction dataset and large language model for finance. In *Advances in Neural Information Processing Systems (Vol. 36)*, pp. 33469-33484. https://papers.nips.cc/paper/2023/hash/6a386d703b50f1cf1f61ab02a15967bb-Abstract-Datasets_and_Benchmarks.html
- [16] Guo, X., Xia, H., Liu, Z., Cao, H., Yang, Z., Liu, Z., Wang, S., Niu, J., Wang, C., Wang, Y., Liang, X., Huang, X., Zhu, B., Wei, Z., Chen, Y., Shen, W., & Zhang, L. (2025). FinEval: A Chinese financial domain knowledge evaluation benchmark for large language models. In *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*. Albuquerque, New Mexico. Association for Computational Linguistics. pp. 6258–6292. doi: 10.18653/v1/2025.naacl-long.318. <https://aclanthology.org/2025.naacl-long.318/>
- [17] Lopez-Lira, A., & Tang, Y. (2023). Can ChatGPT forecast stock price movements? Return predictability and large language models. *SSRN Electronic Journal*. doi: 10.2139/ssrn.4412788. <https://doi.org/10.2139/ssrn.4412788>
- [18] Tao, J., Deokar, A. V., & Deshmukh, A. (2018). Analysing forward-looking statements in initial public offering prospectuses: A text analytics approach. *Journal of Business Analytics*, 1(1), 54–70. doi: 10.1080/2573234X.2018.1507604. <https://doi.org/10.1080/2573234X.2018.1507604>
- [19] Bertrand, M., Duflo, E., & Mullainathan, S. (2004). How much should we trust differences-in-differences estimates? *The Quarterly Journal of Economics*, 119(1), 249–275. doi: 10.1162/003355304772839588. <https://doi.org/10.1162/003355304772839588>