Online Appendix

for

"Public Enforcement and IPO Reporting Quality in Weak Institutional Environments: Evidence from a Random Experiment in China"

Manyi Fan, Kai Guo, and Bin Ke

OA1. Test on randomness

Since our identification strategy is based on the CSRC's random selection of IPO applicants for onsite inspection, we first examine whether the selection is truly random before testing our hypotheses. According to the SAC, the organizer of the random draw, the selection was done by a computer program to ensure randomness. In addition, the SAC also invites investment bank representatives and media to witness the whole process. The literature also shows that regulators may ensure the selection is random to deter low-quality disclosure (Mookherjee and Png 1989). However, the literature has also documented that regulators have severe resource constraints (Kedia and Rajgopal 2011; Ege et al. 2020; Bonsall et al. 2024; Gunny and Hermis 2020), casting doubt on whether the CSRC's random inspection program is truly random. In addition, there might be collusion in the selection process if IPO applicants influence the officials involved.

We examine randomness by comparing selected firms with non-selected firms across the following dimensions: financial performance, internal control quality, ownership structure, intermediaries' reputation, information asymmetry between IPO applicants and regulators, and business processes typically examined by onsite inspectors. We examine financial performance and internal control because they are key considerations for regulators. Ownership structure includes the largest shareholder's ownership and the nature of their ownership. The largest ownership is a key determinant of corporate governance (Fan and Wong 2002; Chen et al. 2006). We also consider state ownership, which reflects securities regulators' preference for state-owned enterprises (SOEs) to go public (Chen et al. 2018; Qian et al. 2024), as well as the differences between SOEs and non-SOEs in earnings management (Chen et al. 2011). Additionally, we take into account the reputation of intermediaries who monitor the IPO applicants. Since an onsite visit can be an information-acquisition activity (Cheng et al. 2016), we examine information asymmetry between IPO applicants and regulators. Finally, we capture characteristics of the five business processes, including revenue, cost, expenses, long-term assets, and financial management processes, that onsite inspection would cover (Shao 2018).

Our financial performance measurements include several metrics. After-tax operating ROE (*ROE*) is one of the CSRC's most closely monitored metrics. We also analyze the gross profit rate (*GROSS*), another commonly used profitability metric. Additionally, we consider leverage (*LEV*), operating revenue (*REVENUE*), and the growth rates of revenue and earnings (*REVENUE_GROWTH*, *EARNING_GROWTH*). To capture the ownership structure, we use *OWNERSHIP_LARGE* to represent the largest shareholder's ownership and *SOE* to indicate whether a firm is controlled by a government agency. We measure intermediaries' reputation using two dummy variables. *TOP10* equals one if the auditor is ranked among the top 10 accounting firms by the Chinese Institute of Certified Public Accountants (CICPA), and *AA_UNDERWRITER* equals one if the underwriter is rated as an AA security firm by the CSRC.[1] Regarding the factors related to internal control quality, in addition to the above-mentioned profitability, ownership structure measurements and auditor's monitoring, we further look at the applicant's involvement in foreign sales (*FOREIGN_RATIO*), size (*SIZE*) and the proportion of cash to total assets (*CASH*) (Ge et al. 2017; Doyle et al. 2007; Ge and McVay 2005; Ashbaugh-Skaife et al. 2009). To capture information asymmetry between the IPO applicant and regulators, we consider size (*SIZE*) and foreign sales involvement (*FOREIGN_RATIO*). Additionally, we include variables such as R&D involvement (*RD_RATIO*), the geographic distance from the firm's headquarters to the CSRC local office (*LN_DISTANCE*), and whether the firm had applied for an IPO before (*MULTIPLE*). Finally, while the above variables have included the characteristics of most of the five business processes that onsite inspection would cover,[2] we add one more variable to capture the proportion of fixed assets (*FIXED_ASSET*). All time-variant variables are constructed using information from one year before the applicant files for the IPO (Li et al. 2021; Miao et al. 2022). Please see Appendix OA for the variable definitions.

Table OA2 shows the comparison results. Columns (2) and (4) show the mean of the non-selected and the selected applications, respectively. Columns (5) and (6) show the differences between those two groups and the corresponding t-statistics, respectively, and Column (7) shows the p-value of the Kolmogorov-Smirnov test (KS test). The null of our KS test is that the selected and non-selected applications are drawn from the same distribution. As shown in Columns (6) and (7), we do not observe

[1] Each year, the CICPA association ranks the top 100 auditing firms in China, mainly based on the total revenue and penalties of the audit firms. Each year, the CSRC also rates each licensed securities firm in China, with AA the de facto highest rate. According to the CSRC, securities firms are evaluated based on their risk management capability, continuous compliance status, and business development status.

[2] We capture the characteristics of revenue process using *REVENUE*, *REVENUE_GROWTH* and *FOREIGN_RATIO*, the characteristics of costing and expenses using *ROE* and *RD_RATIO*, and financial management process using *CASH* and *LEV*.

significant differences between the selected applications and the non-selected ones on any dimension. Hence, there is no evidence that the selection is not random. We also plot the distribution of these characteristics for the selected and non-selected applications in Figure OA2. Consistent with the randomization, the distributions are very similar for selected and non-selected applications.

OA2. The confounding effect of risk-based onsite inspection

The effect of the random onsite inspection program could be confounded by another CSRC rule issued in 2016. In the fourth quarter of 2016, the CSRC introduced a risk-based onsite inspection program for IPO applicants. Specifically, during the IPO paper review stage, if the reviewers raise significant concerns regarding the applicant's compliance with listing or information disclosure requirements and the applicant fails to provide a satisfactory explanation, the applicant may be subject to an onsite inspection.

Unfortunately, because the CSRC does not disclose the identities of firms selected for risk-based onsite inspections, we are unable to exclude them from our control group, potentially contaminating our inferences. However, if the risk-based onsite inspections affect the effectiveness of the random-based onsite inspections, the difference in quality between the randomly selected and non-randomly selected IPO applicants, and therefore the application failure rate between the two groups, should decrease after Q4 2016, because only non-randomly selected firms could be selected on a risk-based basis. To investigate the potential effect of risk-based onsite inspection on our results, we run the following model:

$$FAIL = \beta_0 + \beta_1 SELECTED + \beta_2 SELECTED \times POST2016Q4 + Controls + Industry\ fixed\ effects + Draw\ fixed\ effects + Region\ fixed\ effects \quad \text{(OA2)}$$

where *POST2016Q4* equals one if the random selection is after Q4 2016 (inclusive) and zero otherwise. All the other variables are the same as those in Model (1). *POST2016Q4* is absorbed by the draw fixed effects.

Column (1) in Table OA3 shows our results. The coefficient on *SELECTED* is significantly positive, consistent with the full-sample results in Table 2. As the risk-based onsite inspection program does not contaminate this coefficient, our results in Table 2 are unlikely to be due to the confounding effects of this policy. Consistent with this argument, the coefficient on *SELECTED×POST2016Q4* is not significant. In Columns (2) and (3), we replace the dependent variable with *WITHDRAW* and *VETO* and rerun Model (OA2). Similar to Column (1), the coefficient on *SELECTED×POST2016Q4* is not significant.

OA3. Construction of DACC and PMDACC

DACC is calculated as follows. We estimate the following modified Jones (1991) model cross-sectionally each year using all seasoned public firms in the same two-digit SIC code.[3]

$$TA_{it} = \beta_0 + \beta_1\left(\frac{1}{ASSET_{it-1}}\right) + \beta_2(\Delta SALES_{it} - \Delta AR_{it}) + \beta_3 PPE_{it} + \varepsilon_{it} \quad \text{(OA3\_1)}$$

where TA is total accruals using the indirect cash flow method (i.e., income before extraordinary items minus operating cash flow) scaled by lagged total assets (ASSET), ΔSALES is change in sales scaled by lagged total assets, ΔAR is change in accounts receivable scaled by lagged total assets, and PPE is net property, plant and equipment scaled by lagged total assets.

DACC for any firm i in year t is computed as follows:

$$DACC_{it} = TA_{it} - \widehat{\beta_0} - \widehat{\beta_1}\left(\frac{1}{ASSET_{it-1}}\right) - \widehat{\beta_2}(\Delta SALES_{it} - \Delta AR_{it}) - \widehat{\beta_3} PPE_{it} \quad \text{(OA3\_2)}$$

where $\widehat{\beta_0}, \widehat{\beta_1}, \widehat{\beta_2}, \widehat{\beta_3}$ are the estimates of $\widehat{\beta_0}, \widehat{\beta_1}, \widehat{\beta_2}, \widehat{\beta_3}$ from the last equation.

PMDACC is calculated as the difference in *DACC* between IPO firm *i* in year *t* and a matched listed firm in the same two-digit industry as firm *i* with the closest ROA (defined as net income divided by average total assets) in year *t*. For matching, we require the absolute difference in ROA between firm *i* and its matching firm to be no larger than 20% of firm *i*'s ROA. If we cannot find a matched firm at the two-digit industry level, we use a one-digit industry instead.

Chinese IPO applicants are required to disclose three years of financial statements in the IPO prospectus. Therefore, we can calculate *DACC* and *PMDACC* for the last two years. In the subsequent regressions, we use the average values of *DACC* and *PMDACC* as proxies for *INFO_QUALITY*.

OA4. Long-term implications of the random inspection program

[3] For each two-digit industry year, we require a minimum of 10 firms to estimate the model. If a two-digit industry year has fewer than 10 observations, we use one-digit industry instead.

As an exploratory analysis, we examine whether the random inspection program affects the financial reporting quality of approved IPO firms in the post-IPO period. Section OA4.1 develops the hypotheses, Section OA4.2 outlines the research design, and Section OA4.3 presents and discusses the results.

OA4.1. Hypotheses

We examine whether being randomly selected for onsite inspection affects a firm's financial reporting in the years following the IPO. Consistent with Stein's (1989) rational expectation model (see also Shivakumar (2000) and Chen et al. (2018)), once a firm has selected an optimal (low or high) quality financial reporting strategy at the IPO, rational investors would expect the firm's management to maintain the strategy after the IPO and incorporate the expectation into stock prices. Anticipating that investors hold these beliefs, management that has chosen low-quality reporting at the IPO would continue to manipulate earnings in the post-IPO period because the firm cannot credibly signal the absence of earnings management. Likewise, management that has chosen high-quality reporting at the IPO would continue it in the post-IPO period to retain the benefits of this strategy. Since firms that have been randomly selected for onsite inspection and have obtained IPO approval are expected to have higher-quality pre-IPO financial reports (H3), they should also have higher earnings quality in the post-IPO period. This discussion leads to the following hypothesis:

OA4_H1: Among the IPO applicants that have been approved for listing, applicants randomly selected for onsite inspection are less likely than non-selected applicants to misstate their financial reporting after going public.

We also consider the impact of the random inspection program on the earnings response coefficient (ERC), another indirect proxy for financial reporting quality. Since prior research shows that lower earnings quality is associated with lower ERCs (Francis and Ke 2006), we predict that:

OA4_H2: Among IPO applicants approved for listing, randomly selected applicants for onsite inspection have a higher ERC than non-selected applicants.

OA4.2. Research design

OA4.2.1. Regression model for OA4_H1

To test OA4_H1, we employ the following OLS regression model:

$$MISSTATEMENT\_POST/MISSTATEMENT\_ACC\_POST = \beta_0 + \beta_1 SELECTED + CONTROLS + industry\ fixed\ effects + year - qtr\ fixed\ effects \tag{OA4\_1}$$

The sample firms are those that participated in the random draws after April 2014 and were listed by 2020. The samples are firm quarters from the listing year up to 2020. Our sample period ends in 2020 due to the availability of restatement data, as mentioned in Section 5.2.3. *MISSTATEMENT_POST* equals one if a company's annual report (financial statements or other parts of the report) for a post-IPO year is restated and zero otherwise. Similarly, *MISSTATEMENT_ACC_POST* equals one if a company's financial statements for a post-IPO year are restated and zero otherwise. To maintain a consistent measurement window for the misstatement variables, all restatement announcements must be limited to three years following the fiscal year-end.

We include a set of control variables based on prior literature (e.g., Kinney and McDaniel 1989; Archambeault et al. 2008; Huang and Ke 2018; Lennox and Wu 2022). First, we control for firm size (*SIZE*) because the probability of material errors may be higher for small firms with weaker financial information systems. Next, we control for the firm's financial condition using ROA (*ROA*) and leverage (*LEV*). We also include firm age (*AGE*) because firms' experience in the capital market may affect the probability of restatement. The market-to-book ratio (*MB*) is used to control for the firm's growth opportunities. Finally, we include industry-fixed effects and year-quarter fixed effects. All continuous variables are winsorized at the 1st and 99th percentiles. The standard errors are double-clustered at the firm and year-quarter level. OA4_H1 predicts a negative β_1.

OA4.2.2. Regression model for OA4_H2

To test OA4_H2, we run the following OLS regression model:

$$CAR = \beta_0 + \beta_1 UE + \beta_2 UE * SELECTED + CONTROLS + UE * CONTROLS + firm\ fixed\ effects + year - qtr\ fixed\ effects \tag{OA4\_2}$$

Our sample firms are the approved IPO applicants who participated in the random draws after April 2014 and were listed by 2021. For each firm, our sample contains all firm-quarters from the listing year to 2021. CAR is the market-adjusted 5-day stock return during (-2,2) centered on the earnings announcement day, following Core and Schrand (1999) and Ke and Zhang (2019). *UE* is unexpected earnings, scaled by the stock price two days before the earnings announcement. Unexpected earnings are calculated as the difference between the quarter's EPS and the same quarter last year's EPS, where EPS is calculated using earnings before extraordinary items, adjusted for the effects of share dividends and share splits (Ke and Zhang 2021).[4]

Following the prior literature (Ke and Zhang 2019; Zhu 2019; Chen et al. 2018), we include the following controls. First, we control for size (*SIZE*) as firms of different sizes face different information environments. Second, we include leverage (*LEV*) and daily return volatility (*VOLATILITY*) as firm risks may also affect ERC. We further control growth opportunities (*MTB*) and the absolute value of unexpected earnings (*ABSOLUTE_UE*) to account for the nonlinearity in the ERC. We also control for whether the firm made a loss in the quarter (*LOSS*).

In addition, the regression model includes the interactions of *UE* and controls (Ke and Zhang 2019; Zhu 2019). We further include the interaction of *UE* with the variable *QTR4* to account for differences in ERC between the last fiscal quarter and other quarters, and the interaction of *UE* with *SOE* to control for systematic differences between SOEs and non-SOEs.[5] Firm fixed effects and year-quarter fixed effects are also included in the model. All continuous variables are winsorized at the 1st and 99th percentiles. The standard errors are double-clustered at the firm and year-quarter level. OA4_H2 predicts the coefficient on the interaction term between *UE* and *SELECTED* to be positive.

OA4.3. Regression results

OA4.3.1. Empirical results for OA4_H1

OA4_H1 investigates whether approved IPO firms randomly selected for onsite inspection are less likely to have misstatements after going public. Table OA4.1 shows the results. Panel A shows summary statistics for the variables used in the regression Model (OA4_1). We find that 7% of the

[4] We do not use analyst consensus because most analysts only issue annual forecasts for their target firms.

[5] Please note that variables SELECTED, SOE, and QTR4 are absorbed by the firm fixed effects and year-quarter fixed effects.

observations have restatements and 2% have financial statement-related restatements. Panel B shows summary statistics for non-selected and selected firms separately.

Panel C shows our regression results. In Column (1), we use *MISSTATEMENT_POST* as the dependent variable and find a significantly negative coefficient on *SELECTED* (coefficient = -0.023 with t = -3.08), suggesting that randomly inspected IPO applicants that receive the final IPO approval are 2.3% less likely to restate their post-IPO periodic reports, representing 32.9% of the sample mean. As for controls, we find that firms with higher leverage are more likely to restate, consistent with prior studies (e.g., Lennox and Wu 2022). In Column (2), we use *MISSTATEMENT_ACC_POST* as the dependent variable and continue to find a significant negative coefficient on SELECTED (coefficient = -0.010, t = -2.28). The coefficient magnitude is about 50% of the sample average likelihood of restating accounting information. Overall, the results show that firms randomly selected for onsite inspection are less likely to restate their financial reports after going public, suggesting that onsite inspection is associated with increased financial reporting quality even after the IPO.

OA4.3.2. Empirical results for OA4_H2

In this section, we use ERC as an alternative proxy for disclosure quality, which directly reflects the usefulness of accounting information to investors (Dechow et al. 2010). Table OA4.2 shows the results. Panel A shows summary statistics for the variables used in the regression Model (OA4_2). Panel B of Table OA4.2 shows the regression results. The coefficient on *UE*SELECTED* is positive and significant, suggesting that randomly inspected IPO applicants that receive the final approval enjoy a higher ERC than non-randomly inspected applicants. This finding further suggests that onsite inspection helps weed out IPO applicants with lower reporting quality, thereby leading to higher reporting quality among the approved applicants.

OA5. The CSRC learning hypothesis

As noted in footnote 8 of the paper, if the onsite inspection program serves as a valuable learning opportunity for CSRC inspectors, the financial reporting quality for the subsequently approved IPO applicants should be higher, even if they are not selected by the random inspection program. Because the benefit of this learning effect should be largest for subsequent IPO applicants, as with randomly selected IPO applicants, we operationalize this hypothesis by focusing on those in the same industry or province as the randomly selected IPO applicants.

To test this hypothesis, we adopt the following regression model:

$$MISSTATEMENT\_\mathrm{IPO} = \beta_0 + \beta_1 CUL\_SELECTED\_IND + \beta_2 CUL\_SELECTED\_REGION + CONTROLS + industry\ fixed\ effects + draw\ fixed\ effects + region\ fixed\ effects \tag{OA5}$$

MISSTATEMENT_IPO is as previously defined. *CONTROLS* are the same as those in Model (4) of the paper. The key independent variables of interest are $CUL\_SELECTED\_IND$ and $CUL\_SELECTED\_REGION$. $CUL\_SELECTED\_IND$ is defined as the number of IPO applicants from the same industry as the focal applicants who were selected for the random inspection program before the focal applicants submitted their IPO materials. $CUL\_SELECTED\_REGION$ is defined similarly for IPO applicants from the same province. Inferences are similar if we define both variables using dummies (untabulated). The CSRC learning hypothesis predicts the coefficients on both $CUL\_SELECTED\_IND$ and $CUL\_SELECTED\_REGION$ to be negative.

The sample for Model (OA5) includes all IPO applications for the Shanghai and Shenzhen stock exchanges from 2014 to 2020. We exclude IPO applicants selected for onsite inspection because the learning hypothesis focuses on non-selected applicants. We limit IPO applicants under the approval-based system because our focus is on CSRC staff learning.

The regression results are shown in Table OA5.1. The coefficients on $CUL\_SELECTED\_IND$ and $CUL\_SELECTED\_REGION$ are never significantly negative, inconsistent with the CSRC learning hypothesis.

To further rule out the CSRC learning hypothesis, we examine whether the deterrence effect in Model (5) differs between the first half (*POST1*) and the second half (*POST2*) of the post-program period. If the learning hypothesis is valid, we should expect the deterrence effect to be larger in the second half of the post-program period. Table OA5.2 reports the results of this test. We find no evidence supporting this prediction.

Appendix OA. References

Archambeault, D. S., F. T. DeZoort, and D. R. Hermanson. 2008. Audit committee incentive compensation and accounting restatements. *Contemporary Accounting Research* 25 (4):965-992.

Ashbaugh-Skaife, H., D. W. Collins, W. R. Kinney Jr, and R. LaFond. 2009. The effect of SOX internal control deficiencies on firm risk and cost of equity. *Journal of Accounting Research* 47 (1):1-43.

Baghai, R. P., R. C. Silva, and L. Ye. 2024. Teams and bankruptcy. *The Review of Financial Studies 37*(9): 2855-2902.

Bonsall, S. B., E. R. Holzman, and B. P. Miller. 2024. Wearing out the watchdog: The impact of SEC case backlog on the formal investigation process. *The Accounting Review* 99 (1):81-104.

Brown, S. V., and J. W. Tucker. 2011. Large-sample evidence on firms' year-over-year MD&A modifications. *Journal of Accounting Research 49*(2): 309-346.

Chen, G., M. Firth, D. N. Gao, and O. M. Rui. 2006. Ownership structure, corporate governance, and fraud: Evidence from China. *Journal of Corporate Finance* 12 (3):424-448.

Chen, H., J. Z. Chen, G. J. Lobo, and Y. Wang. 2011. Effects of audit quality on earnings management and cost of equity capital: Evidence from China. *Contemporary Accounting Research* 28 (3):892-925.

Chen, J., B. Ke, D. Wu, and Z. Yang. 2018. The consequences of shifting the IPO offer pricing power from securities regulators to market participants in weak institutional environments: evidence from China. *Journal of Corporate Finance* 50:349-370.

Cheng, Q., F. Du, X. Wang, and Y. Wang. 2016. Seeing is believing: Analysts' corporate site visits. *Review of Accounting Studies* 21 (4):1245-1286.

Core, J.E., and C.M. Schrand. 1999. The effect of accounting-based debt covenants on equity valuation. *Journal of Accounting and Economics 27*(1): 1-34.

Dechow, P., W. Ge, and C. Schrand. 2010. Understanding earnings quality: A review of the proxies, their determinants and their consequences. *Journal of Accounting and Economics* 50 (2-3):344-401.

Doyle, J., W. Ge, and S. McVay. 2007. Determinants of weaknesses in internal control over financial reporting. *Journal of Accounting and Economics* 44 (1-2):193-223.

Ege, M., J. L. Glenn, and J. R. Robinson. 2020. Unexpected SEC resource constraints and comment letter quality. *Contemporary Accounting Research* 37 (1):33-67.

Fan, J. P., and T. J. Wong. 2002. Corporate ownership structure and the informativeness of accounting earnings in East Asia. *Journal of Accounting and Economics* 33 (3):401-425.

Francis, J. R., and B. Ke. 2006. Disclosure of fees paid to auditors and the market valuation of earnings surprises. *Review of Accounting Studies* 11:495-523.

Ge, W., A. Koester, and S. McVay. 2017. Benefits and costs of Sarbanes-Oxley Section 404 (b) exemption: Evidence from small firms' internal control disclosures. *Journal of Accounting and Economics* 63 (2-3):358-384.

Ge, W., and S. McVay. 2005. The disclosure of material weaknesses in internal control after the Sarbanes-Oxley Act. *Accounting Horizons* 19 (3):137-158.

Gunny, K. A., and J. M. Hermis. 2020. How busyness influences SEC compliance activities: Evidence from the filing review process and comment letters. *Contemporary Accounting Research* 37 (1):7-32.

Huang, J., and B. Ke. 2018. The consequences of shifting corporate disclosure enforcement from public to private in weak institutional environments: Are market institutions ready. *Shanghai University of Finance and Economics working paper*.

Ke, B., and X. Zhang. 2019. The consequences of management earnings forecast regulation: stock price informativeness and disclosure externalities. *Available at SSRN 3327660*.

———. 2021. Does public enforcement work in weak investor protection countries? Evidence from China. *Contemporary Accounting Research* 38 (2):1231-1273.

Kedia, S., and S. Rajgopal. 2011. Do the SEC's enforcement preferences affect corporate misconduct? *Journal of Accounting and Economics* 51 (3):259-278.

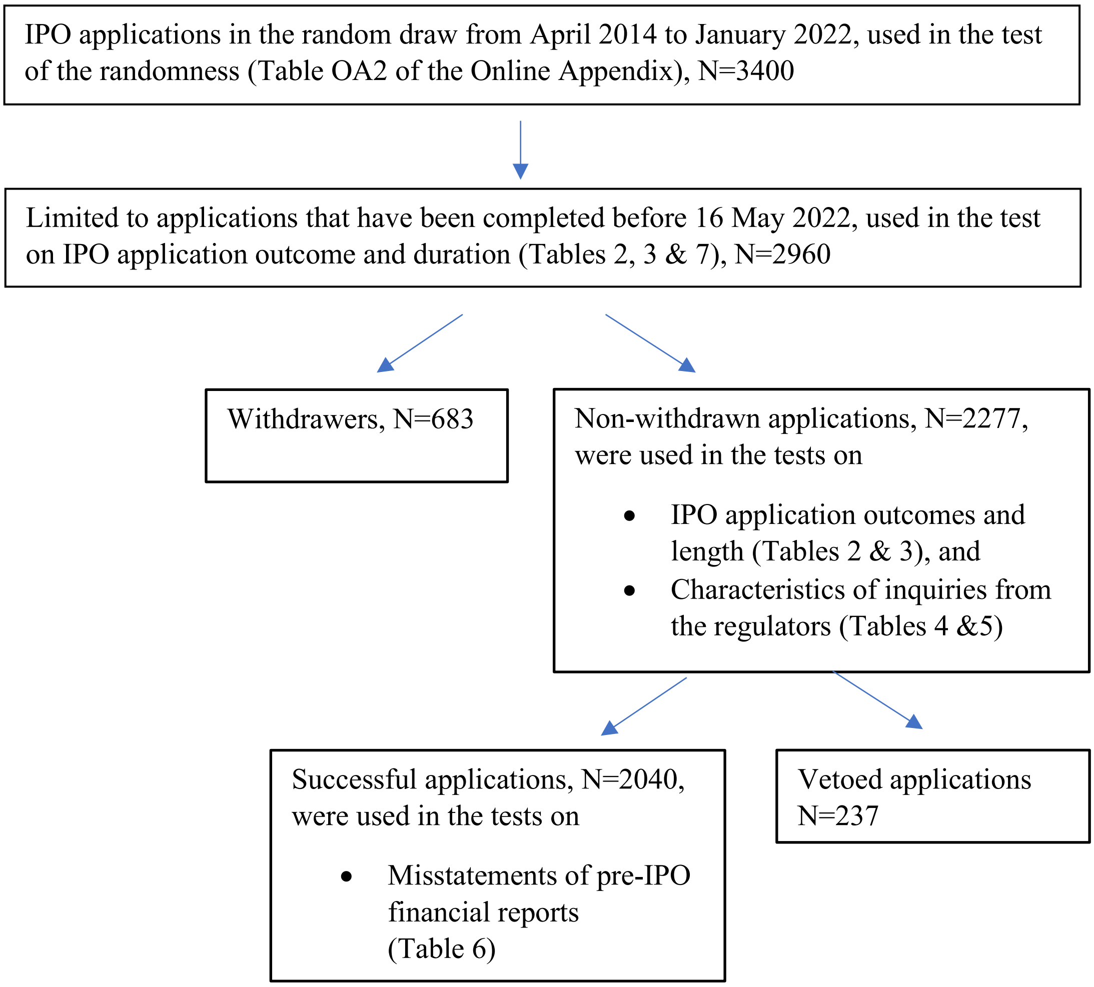
Kinney Jr, W. R., and L. S. McDaniel. 1989. Characteristics of firms correcting previously reported quarterly earnings. *Journal of Accounting and Economics* 11 (1):71-93.
Lennox, C. S., and X. Wu. 2022. Mandatory internal control audits, audit adjustments, and financial reporting quality: Evidence from China. *The Accounting Review* 97 (1):341-364.
Li, Y., H. Shi, and Y. Zhou. 2021. The influence of the media on government decisions: Evidence from IPOs in China. *Journal of Corporate Finance* 70:102056.
Miao, M., W. Wang, and Z. Zuo. 2022. Polluted IPOs. *Available at SSRN 4059255*.
Mookherjee, D., and I. Png. 1989. Optimal auditing, insurance, and redistribution. *The Quarterly Journal of Economics* 104 (2):399-415.
Qian, Y., J. R. Ritter, and X. Shao. 2024. Initial public offerings Chinese style. *Journal of Financial and Quantitative Analysis* 59 (1):1-38.
Shao J. 2018. The impact of the onsite inspection of IPO companies by the China Securities Regulatory Commission on companies planning to list (in Chinese). *Commercial Accounting* 5 (10):17-19.
Shivakumar, L. 2000. Do firms mislead investors by overstating earnings before seasoned equity offerings? *Journal of Accounting and Economics* 29 (3):339-371.
Stein, J. C. 1989. Efficient capital markets, inefficient firms: A model of myopic corporate behavior. *The Quarterly Journal of Economics* 104 (4):655-669.
Zhu, C. 2019. Big data as a governance mechanism. *The Review of Financial Studies* 32 (5):2021-2061.

Appendix OA. Variable definition

Variable	Definition
ROE	After-tax operating ROE, calculated as the net profit after deducting non-recurring profit and loss items over book equity
GROSS	Gross profit rate
SOE	A dummy variable that equals one if a firm's ultimate owner is a government agency
SIZE	Natural logarithm of total assets
FOREIGN_RATIO	Overseas revenue over the total revenue
RD_RATIO	R&D expense over the total revenue
LN_DISTANCE	Natural logarithm of the distance from the headquarters of the firm to the CSRC local office in charge of the firm
MULTIPLE	A dummy variable that equals one if the firm has applied for an IPO before a specific selection
TOP10	A dummy variable that equals one if the audit firm is a top 10 firm rated by the CICPA one year before the selection
AA_UNDERWRITER	A dummy variable that equals one if the underwriter is rated as an AA security firm by the CSRC one year before the selection
REVENUE	The natural logarithm of total revenue
REVENUE_GROWTH	Growth rate of annual revenue
LEV	Firm leverage ratio
FIXED_ASSET	Fixed asset divided by total assets
OWNERSHIP_LARGE	The percentage of ownership of the largest shareholder
CASH	Cash divided by total assets
EARNING_GROWTH	Growth rate of net profits
SELECTED	A dummy variable that equals one if the firm is selected for random onsite inspection and zero otherwise
FAIL	A dummy variable that equals one if the application is withdrawn or vetoed, and zero otherwise
WITHDRAW	A dummy variable that equals one if the application is withdrawn by the applicant and zero otherwise
VETO	A dummy variable that equals one if the IPO application is vetoed and zero otherwise
POST2016Q4	A dummy variable that equals one if the selection is after Q4 2016 (inclusive) and zero otherwise
REGISTRATION	A dummy variable that equals one if the IPO application is under the registration-based IPO system and zero otherwise
DACC	The Modified Jones Model discretionary accruals. See Online Appendix OA3 for more details.
PMDACC	Signed performance-matched modified Jones Model discretionary accruals. See Online Appendix OA3 for more details.
UE	Unexpected earnings, scaled by the stock price two days before the earnings announcement. Unexpected earnings are calculated as the difference between the EPS of the fiscal quarter and the EPS of the same quarter last year. The EPS is based on earnings before extraordinary items

	(Ke and Zhang 2021). When calculating the unexpected earnings, we also adjust for the effect of share dividends and share splits on EPS.
VOLATILITY	Daily return volatility, calculated as the standard deviation of stock return during the past 90 days, before two days before the earnings announcement.
MTB	Market-to-book ratio
ABSOLUTE_UE	The absolute value of unexpected earnings
QTR4	A dummy variable that equals one if it is the 4th quarter.
LOSS	A dummy variable that equals one if the firm makes losses in the quarter
CUL_SELECTED_IND	The number of IPO applicants from the same industry as the focal applicants who were selected for the random inspection program before the focal applicants submitted their IPO materials.
CUL_SELECTED_REGION	The number of IPO applicants from the same region as the focal applicants who were selected for the random inspection program before the focal applicants submitted their IPO materials.
MISSTATEMENT_POST	A dummy variable equal to one if the annual reports for a post-IPO year are restated within three years following the fiscal year-end, and zero otherwise
MISSTATEMENT_ACC_POST	A dummy variable equal to one if the financial statements for a post-IPO year are restated within three years following the fiscal year-end, and zero otherwise
AGE	Number of years since a firm was listed.
ROA	After-tax operating ROA, calculated as the net profit after deducting non-recurring profit and loss items over total assets
MB	Market-to-book ratio
CAR	Market-adjusted 5-day stock return during (-2,2), where day 0 is the earnings announcement day.
STARBOARD	A dummy variable that equals one if the IPO application is for the STAR Board and zero otherwise
POST1	A dummy variable that equals one if the IPO application is submitted during the first half of the period from October 2020 to the end of 2022, and zero otherwise
POST2	A dummy variable that equals one if the IPO application is submitted during the second half of the period from October 2020 to the end of 2022, and zero otherwise

Figure OA1. Sample selection for the tests on the direct effects of the random inspection program[6]



[6] The sample size in different tests may vary due to the data availability. Please refer to the details in each table for further information.

Figure OA2. Distribution of characteristics of IPO applications in selected and non-selected groups

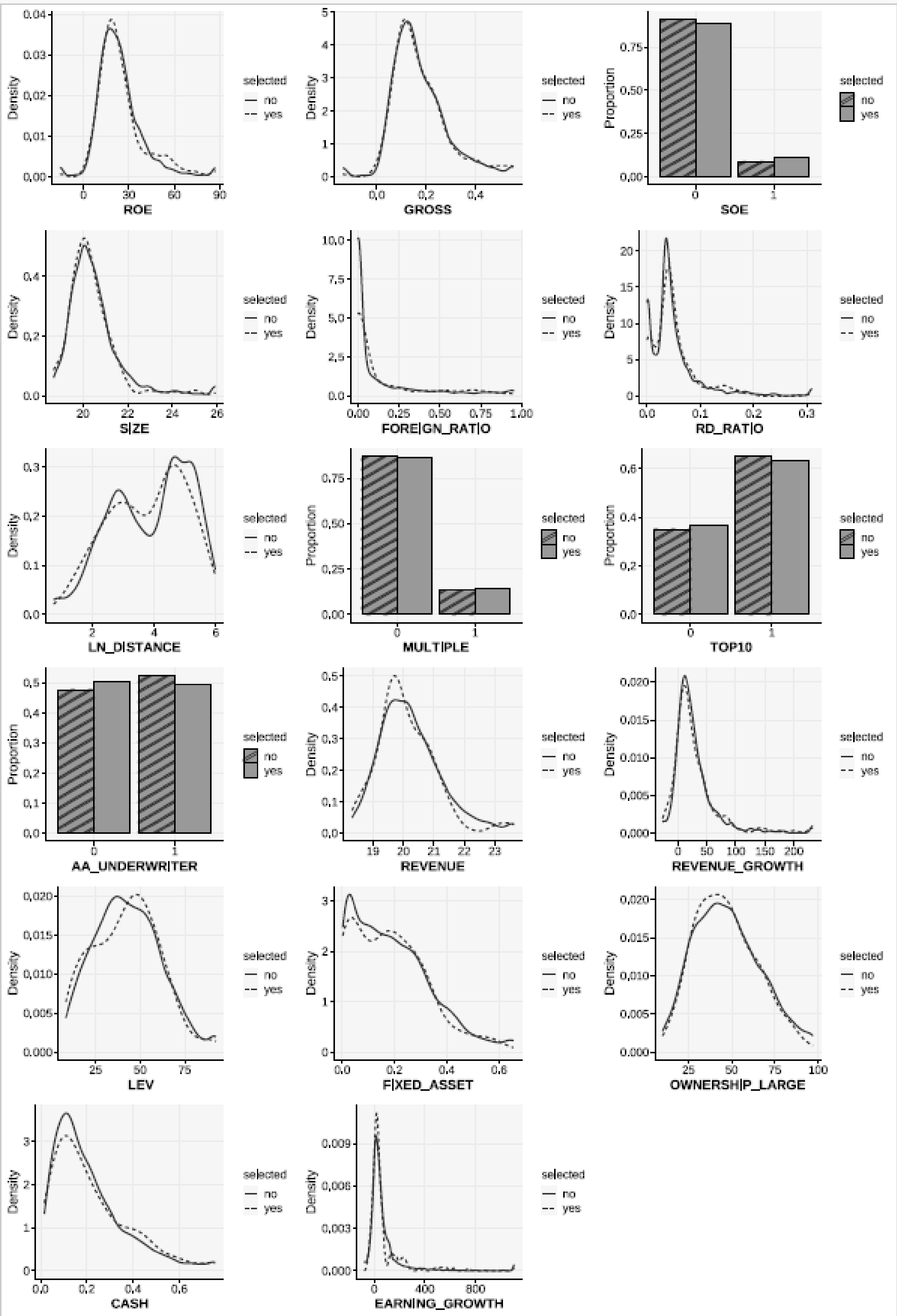


Table OA1. Distribution of the size of the random draw pool and the number of selected firms for onsite inspection

The table shows the number of firms in the selection pool and the number selected for onsite inspection.

Date	No. of firms in the selection pool[7]	No. of selected firms
25 Apr 2014	20	1
16 May 2014	20	1
20 Jun 2014	20	1
8 Aug 2014	20	1
5 Sep 2014	20	1
10 Oct 2014	20	1
14 Nov 2014	20	1
28 Nov 2014	40	2
12 Dec 2014	40	2
5 Mar 2015	40	2
13 Mar 2015	39	2
3 Apr 2015	78	4
10 Apr 2015	79	4
24 Apr 2015	279	14
29 Jul 2016	99	5
2 Dec 2016	198	10
24 Feb 2017	120	6
28 Apr 2017	120	6
28 Jul 2017	200	10
1 Dec 2017	198	10
25 May 2018	65	3
16 Nov 2018	81	4
30 Apr 2019	99	5
11 Jul 2019	237	11
13 Mar 2020	74	4
25 Aug 2020	80	4
14 Jan 2021	29	2
31 Jan 2021	407	20
25 Apr 2021	32	2
4 Jul 2021	377	19
10 Oct 2021	63	3
7 Jan 2022	186	9
Total	3400	170

[7] We infer the identity of firms included in each selection pool from the CSRC IPO application status weekly reports. Please see Section 2.2 for more details.

Table OA2. Differences between selected firms and non-selected firms in all random draws

This table shows the differences between selected firms and non-selected firms in all random draws. The sample includes all IPO applicants who participated in the random draws from April 2014 to January 2022. Please see Appendix OA for the variable definitions. ***, **, and * indicate that the difference between the non-selected and selected firms in the corresponding variable is statistically significant at the 0.01, 0.05, and 0.10 levels, respectively.

	(1) N	(2) Mean	(3) N	(4) Mean	(5) mean-diff	(6) t-value	(7) p-value of Kolmogorov-Smirnov Test
	Non-selected		Selected				
ROE	3206	23.730	169	25.037	-1.307	-1.099	0.703
GROSS	3190	0.174	169	0.179	-0.005	-0.577	0.729
SOE	3229	0.089	170	0.112	-0.023	-1.016	1.000
SIZE	3212	20.509	169	20.379	0.129	1.363	0.305
FOREIGN_RATIO	3204	0.143	169	0.150	-0.006	-0.331	0.982
RD_RATIO	3204	0.049	169	0.049	0.000	0.004	0.116
LN_DISTANCE	3219	3.887	170	3.792	0.095	0.942	0.248
MULTIPLE	3229	0.133	170	0.141	-0.008	-0.287	1.000
TOP10	3229	0.648	170	0.635	0.013	0.351	1.000
AA_UNDERWRITER	3229	0.524	170	0.494	0.030	0.752	0.999
REVENUE	3204	20.238	169	20.104	0.134	1.630	0.140
REVENUE_GROWTH	2935	28.216	156	29.040	-0.824	-0.264	0.649
LEV	3209	43.022	169	42.160	0.862	0.591	0.805
FIXED_ASSET	3201	0.192	168	0.181	0.012	0.971	0.517
OWNERSHIP_LARGE	2917	47.664	144	46.690	0.974	0.589	0.854
CASH	3158	0.212	164	0.220	-0.008	-0.642	0.717
EARNING_GROWTH	2927	74.565	154	75.776	-1.211	-0.091	0.295

Table OA3. The effect of risk-based inspection on Table 2's inference

This table presents the OLS regression results of the following model:

$$OUTCOME = \beta_0 + \beta_1 SELECTED + \beta_2 SELECTED \times POST2016Q4 + CONTROLS \\ + industry\ fixed\ effects + draw\ fixed\ effects + region\ fixed\ effects$$

Please see Appendix OA for the variable definitions. The sample used for Columns (1) and (2) includes all IPO applicants who participated in the random draw from April 2014 to January 2022 and have completed the inspection before 16 May 2022. In Column (3), we remove applications withdrawn before the Issuance Examination Committee Meeting. All continuous variables are winsorized at the 1st and 99th percentiles. The t-values in brackets are based on standard errors adjusted for industry clustering. ***, **, and * indicate that the coefficients are significant at the 0.01, 0.05, and 0.10 levels, based on two-tailed statistical tests.

	(1) FAIL	(2) WITHDRAW	(3) VETO
SELECTED	0.296***	0.210***	0.163**
	(3.62)	(2.76)	(2.47)
SELECTED× POST2016Q4	0.019	0.111	-0.091
	(0.20)	(1.21)	(-1.21)
ROE	-0.003***	-0.002**	-0.002***
	(-3.41)	(-2.48)	(-3.00)
GROSS	-0.066	-0.084	-0.031
	(-0.62)	(-1.01)	(-0.29)
SIZE	-0.091***	-0.064***	-0.053***
	(-6.94)	(-4.84)	(-6.99)
RD_RATIO	-0.677***	-0.516**	-0.327***
	(-2.83)	(-2.42)	(-3.27)
FOREIGN_RATIO	-0.112***	-0.088***	-0.043
	(-3.55)	(-3.20)	(-1.66)
LN_DISTANCE	0.002	0.001	0.001
	(0.26)	(0.12)	(0.26)
TOP10	0.007	0.010	-0.005
	(0.46)	(0.68)	(-0.38)
AA_UNDERWRITER	-0.016	-0.012	-0.005
	(-0.91)	(-0.79)	(-0.28)
SOE	0.016	0.004	0.018
	(0.49)	(0.11)	(0.68)
REGISTRATION	-0.065	-0.027	-0.054**
	(-1.46)	(-0.74)	(-2.22)
N	2906	2906	2257
adj. R^2	0.182	0.141	0.119
Industry	Yes	Yes	Yes
Region	Yes	Yes	Yes
Draw date	Yes	Yes	Yes

Table OA4.1. The impact of onsite inspection on misstatement after firms go public

This table presents the OLS regression results of the likelihood of misstatement on the random selection indicator and control variables after firms go public.

$$MISSTATEMENT\_POST\ /MISSTATEMENT\_ACC\_POST = \beta_0 + \beta_1 SELECTED + CONTROLS + industry\ fixed\ effects + year - qtr\ fixed\ effects + region\ \ fixed\ effect$$

Please see Appendix OA for the variable definitions. Panel A shows the summary statistics used in the regression model. Panel B shows statistics for selected and non-selected applications separately. Panel C shows the regression results. Our sample firms are those that participated in the random draws after April 2014 and were listed by 2020. The samples are firm quarters from the listing to 2020. All continuous variables are winsorized at the 1st and 99th percentiles. The t-values in brackets are based on standard errors adjusted for firm and year-quarter double clustering. ***, **, and * indicate that the coefficients are significant at the 0.01, 0.05, and 0.10 levels, based on two-tailed statistical tests.

Panel A: Summary statistics

	N	mean	sd	p25	median	p75
MISSTATEMENT_POST	17208	0.07	0.25	0.00	0.00	0.00
MISSTATEMENT_ACC_POST	17208	0.02	0.15	0.00	0.00	0.00
SELECTED	17208	0.03	0.17	0.00	0.00	0.00
SIZE	17208	21.42	0.91	20.76	21.28	21.93
LEV	17208	0.33	0.17	0.19	0.31	0.45
AGE	17208	2.13	1.47	1.00	2.00	3.00
ROA	17208	0.02	0.02	0.01	0.01	0.02
MB	17208	4.85	3.16	2.66	3.96	6.05

Panel B. Summary statistics of non-selected and selected applications

	N	Mean	N	Mean	mean-diff	t-value
	Non-selected		Selected			
MISSTATEMENT_POST	16703	0.067	505	0.046	0.021*	1.871
MISSTATEMENT_ACC_POST	16703	0.024	505	0.014	0.010	1.435
SIZE	16703	21.417	505	21.638	-0.221***	-5.359
LEV	16703	0.331	505	0.340	-0.009	-1.108
AGE	16703	2.134	505	2.002	0.132**	1.982
ROA	16703	0.016	505	0.016	0.000	0.661
MB	16703	4.852	505	4.628	0.225	1.576

Panel C: Regression results

	(1) MISSTATEMENT_POST	(2) MISSTATEMENT_ACC_POST
SELECTED	-0.023***	-0.010**
	(-3.08)	(-2.28)
SIZE	-0.005	-0.004
	(-1.31)	(-1.31)
LEV	0.086***	0.057***
	(4.13)	(3.23)
AGE	0.003	0.004**
	(1.01)	(2.62)
ROA	-0.485*	-0.248*
	(-2.05)	(-2.06)
MB	-0.000	-0.000
	(-0.01)	(-0.01)
N	17208	17208
adj. R^2	0.038	0.025
Industry	Yes	Yes
Region	Yes	Yes
Year-qtr	Yes	Yes

Table OA4.2. The impact of onsite inspection on ERC after firms go public

This table presents the OLS regression results of the ERC model.

$$CAR = \beta_0 + \beta_1 UE + \beta_2 UE \times SELECTED + CONTROLS + UE * CONTROLS \\ + firm\ fixed\ effects + year - qtr\ fixed\ effects$$

Please see Appendix OA for the variable definitions. Panel A shows the summary statistics used in the regression model. Panel B shows the regression results. Our sample firms are those that participated in the random draws after April 2014 and were listed by 2021. The samples are firm quarters from the listing to 2021. All continuous variables are winsorized at the 1st and 99th percentiles. The t-values in brackets are based on standard errors adjusted for firm and year-quarter double clustering. ***, **, and * indicate that the coefficients are significant at the 0.01, 0.05, and 0.10 levels, based on two-tailed statistical tests.

Panel A: summary statistics

	N	mean	sd	p25	median	p75
CAR	17577	-0.00	0.07	-0.04	-0.01	0.03
UE	17577	-0.00	0.01	-0.00	0.00	0.00
SELECTED	17577	0.03	0.17	0.00	0.00	0.00
SIZE	17577	21.71	1.15	20.94	21.49	22.19
VOLATILITY	17577	0.03	0.01	0.02	0.03	0.03
LEV	17577	0.37	0.19	0.22	0.35	0.50
MTB	17577	4.10	2.95	2.18	3.24	5.00
LOSS	17577	0.45	0.50	0.00	0.00	1.00
ABSOLUTE_UE	17577	0.01	0.01	0.00	0.00	0.01
SOE	17577	0.10	0.29	0.00	0.00	0.00

Panel B: OLS results

	(1) CAR
UE	1.996
	(1.50)
UE*SELECTED	0.172**
	(2.11)
SIZE	-0.006
	(-1.34)
UE*SIZE	0.038
	(0.67)
VOLATILITY	-0.195
	(-1.37)
UE* VOLATILITY	-7.922
	(-0.97)
LEV	0.011
	(1.03)
UE*LEV	-1.507***
	(-3.82)
MTB	-0.002**
	(-2.32)
UE*MTB	0.016
	(0.27)
LOSS	-0.008***
	(-4.27)
UE*LOSS	0.003***
	(4.12)
ABSOLUTE_UE	-0.126
	(-1.19)
UE* ABSOLUTE_UE	-35.794***
	(-6.84)
UE*SOE	-0.304**
	(-2.64)
UE*QTR4	-0.001
	(-0.97)
N	17,577
adj. R^2	0.100
Firm fixed effects	Yes
Year-qtr fixed effects	Yes

Table OA5.1. Testing the CSRC learning hypothesis

We use the following model to test the CSRC learning hypothesis:

$$MISSTATEMENT\_\text{IPO} = \beta_0 + \beta_1 CUL\_SELECTED\_IND + \beta_2 CUL\_SELECTED\_REGION + CONTROLS + industry\ fixed\ effects + draw\ fixed\ effects + region\ fixed\ effects$$

The independent variable $CUL\_SELECTED\_ind$ ($CUL\_SELECTED\_region$) is defined as the number of IPO applicants that belong to the same industry (or province) as the focal applicants who were selected by the random inspection program before the focal applicants submitted their IPO materials.

The sample includes all IPO applications from 2014 to 2020. We exclude IPO applicants selected by the random inspection program. We also delete IPO applications in the registration system because these firms are subject to the respective stock exchanges' inspection programs rather than the CSRC.

	MISSTATEMENT_IPO
$CUL\_SELECTED\_IND$	-0.001
	(-0.93)
$CUL\_SELECTED\_REGION$	0.001
	(1.42)
CONTROLS	YES
N	1288
adj. R^2	0.118
Industry	YES
Region	YES
Draw date	YES

Table OA5.2. Testing the CSRC learning hypothesis using the deterrence Model (5)

This table tests whether the deterrence effect of the random inspection program differs between the first half (POST1) and the second half (POST2) of the post-random-inspection period. We follow Model (5) in the paper. The dependent variable is INFO_QUALITY, captured by DACC/PMDACC. Please see Appendix A for the variable definitions. Columns (1) and (2) show the result using the full sample, while Columns (3) and (4) show the result using the PSM matched sample. The sample includes IPO applications submitted to the STAR Board and Main Board from 2019 to 2022. All continuous variables are winsorized at the 1st and 99th percentiles. The t-values in brackets are based on standard errors adjusted for industry clustering. ***, **, and * indicate that the coefficients are significant at the 0.01, 0.05, and 0.10 levels, based on two-tailed statistical tests.

	(1)	(2)	(3)	(4)
	Full sample		PSM sample	
	PMDACC	DACC	PMDACC	DACC
STARBOARD	0.023	0.011	0.020	0.007
	(1.12)	(0.60)	(0.68)	(0.38)
STARBOARD×POST1	-0.057**	-0.076***	-0.117**	-0.097*
	(-2.30)	(-3.41)	(-2.33)	(-1.96)
STARBOARD×POST2	-0.024	-0.030	-0.072	-0.065*
	(-1.22)	(-1.56)	(-1.70)	(-2.06)
N	1488	1495	280	280
adj. R^2	0.170	0.152	0.036	0.049
Application year-month	Yes	Yes	Yes	Yes
Industry	Yes	Yes	Yes	Yes
p value for STARBOARD×POST1 vs STARBOARD×POST2	0.1*	0.02**	0.36	0.54

Table OA6. Parallel trends assumption test of H4

This table presents the parallel trends assumption test for H4.[8] The dependent variable is INFO_QUALITY, proxied by DACC or PMDACC. Please see Appendix A for the variable definitions. Columns (1) and (2) show the results using Table 8's sample under PSM. The sample includes IPO applications submitted to the STAR Board and the Main Board from 2019 to 2022. For the pre-period, we classify applications into bins by application quarter and define the event time relative to 2020 Q4, when the random inspection policy was introduced. PRE1, PRE2, PRE3, and PRE4 represent indicator variables for event time quarters –1, –2, –3, and ⩽ –4, relative to 2020 Q4, respectively. Distant leads are collapsed into a single category (PRE4=1) to conserve degrees of freedom and focus on the quarters most relevant to the policy window (e.g., Baghai et al. 2024). The event quarter (2020Q4) is omitted from the sample because the policy was announced mid-quarter. We use PRE1 as the benchmark. The t-values in brackets are based on standard errors adjusted for industry clustering. ***, **, and * indicate that the coefficients are significant at the 0.01, 0.05, and 0.10 levels, based on two-tailed statistical tests.

	PSM sample	
	(1)	(2)
	PMDACC	DACC
STARBOARD×PRE4	-0.013	-0.034
	(-0.26)	(-0.75)
STARBOARD×PRE3	-0.053	0.046
	(-0.33)	(0.35)
STARBOARD×PRE2	-0.056	-0.047
	(-1.03)	(-1.17)
STARBOARD×POST	-0.093*	-0.075*
	(-1.90)	(-2.11)
STARBOARD	0.045	0.023
	(0.80)	(0.58)
SIZE	-0.007	-0.011*
	(-1.07)	(-1.80)
LEV	0.000	-0.000
	(0.48)	(-0.30)
OCF	-0.252***	-0.237**
	(-2.97)	(-2.76)
CURRENT	0.001	-0.004
	(0.13)	(-1.06)
SOE	0.009	-0.003
	(0.33)	(-0.12)
RD_RATIO	0.005	0.002
	(1.17)	(0.94)
N	263	263
adj. R^2	0.084	0.096
Application QTR FE	Yes	Yes
Industry FE	Yes	Yes

[8] The parallel trends test is typically applied in difference-in-differences settings with repeated observations for the same firm over time. In our setting, however, each IPO applicant appears only once in the sample used in Table 8. Strictly speaking, therefore, the standard parallel trends test is not applicable.

Table OA7. Changes of disclosure, governance, and operation between the first and second IPO prospectuses for SELECTED=0 versus SELECTED=1 firms

This table compares changes in disclosure, governance, and operations between the first and second IPO prospectuses for SELECTED=0 and SELECTED=1 firms. The definition of the outcomes is:

dacc_diff and pmdacc_diff: the change in DACC and PMDACC between two IPO applications.

Change owner: a dummy variable that equals 1 if the applicant changed the identity of its controlling owner in the 2nd application.

Change of voting right: the controlling owner's voting right in the 2nd application minus that of the 1st application.

Change of independent directors %: the ratio of the number of independent directors relative to board size in the 2nd application minus that of the 1st application.

% of new independent directors: percentage of new independent directors in the 2nd application over the total number of independent directors in the 2nd application, where new independent directors are defined as those not in the 1st application.

Change in Length: Percentage change of the log of the length of the disclosure of operational information from the 1st to the 2nd application.

Cosine Distance: The cosine distance of the two applications of the same applicant, adjusted for document length, following Brown and Tucker (2011). The raw cosine distance is calculated using the operation disclosure in each application as the corpus. Then we use the TF-IDF (term-frequency and inverse document frequency) weighting to create sparse vectors. Next, we compute the cosine similarity and use 1-cosine similarity as the cosine distance. The raw score ranges from 0 to 1, with a higher score indicating greater dissimilarity. We then adjust the raw cosine distance by the document length.

Change in compliance score: The difference in compliance score over the same applicants' two applications. We assess the extent to which an IPO applicant's prospectus meets CSRC requirements for operational transparency. To do so, we develop a rubric by decomposing the operational disclosure standard into sub-requirements. For each sub-requirement, we prompt an LLM to assign a score of 0 (no information provided), 1 (some information provided), or 2 (the information provided meets the requirement). The total score for each prospectus is the sum of sub-scores divided by the maximum possible score, then multiplied by 100 to convert to a 100-point scale. The score is evaluated by DeepSeek-V3.2-Exp (Thinking Mode) from 20 to 22 October 2025, with the temperature set to 0.

	Obs	Mean	Obs	Mean	Mean Diff.	t value
	SELECTED=0		SELECTED=1			
Disclosure:						
dacc_diff	184	-0.001	22	-0.042	0.041	1.084
pmdacc _diff	179	-0.010	22	-0.064	0.054	1.406
Corporate governance:						
Change Owner	188	0.101	20	0.150	-0.049	-0.577
Change of Voting Right	188	-0.006	20	0.001	-0.007	-0.882
Change of independent directors %	188	-0.002	20	-0.005	0.003	0.616

% of new independent directors	188	0.305	20	0.300	0.005	0.057
Operation disclosure:						
Change in Length	188	0.017	20	0.023	-0.006	-1.121
Cosine Distance (adjusted for length)	188	0.0008	20	0.0046	-0.0038	-0.244
Change in compliance score[9]	188	5.299	20	5.377	-0.078	-0.031

[9] There are variations in disclosure requirements across different boards and one applicant may apply for IPO in different boards in the two attempts. Therefore, the change in disclosure behavior may be caused by the different standard in two applications. To account for this possibility, we also run a regression for robustness using the difference in compliance scores between an applicant's 2nd and 1st applications as the dependent variable. Our key independent variable is SELECTED, with controls for disclosure standards in the 1st and 2nd applications, plus their interaction. The coefficient on SELECTED is insignificant (p=0.64), showing insignificant differences in the change of compliance to the disclosure requirements on operations between the two groups (untabulated).