

Complex Firms, Hidden Signals: Evidence from Earnings Conference Calls

Hirindu Kawshala

Department of Finance and Accounting
Surrey Business School, University of Surrey
Guildford, Surrey, GU2 7XH, United Kingdom
Email: hk00775@surrey.ac.uk

Shuhui Wang*

Department of Finance and Accounting
Surrey Business School, University of Surrey
Guildford, Surrey, GU2 7XH, United Kingdom
Email: shuhui.wang@surrey.ac.uk

* Corresponding Author: Shuhui Wang

ABSTRACT

We use earnings conference calls to capture firm complexity and examine its impact on information asymmetry for U.S. public firms from 2002 to 2023. Our results show that more complex firms exhibit higher information asymmetry, as measured by return volatility and bid–ask spreads, and this relationship persists after controlling for readability and alternative measures of complexity. We further validate the robustness of our findings by employing a residual complexity measure, analyzing within-firm variation over time, and estimating system GMM models. In addition, we exploit the Global Financial Crisis as a period of heightened scrutiny and pressure for transparency, during which firms reduced their complexity and the link between complexity and information asymmetry weakened. Overall, our evidence suggests that firm complexity plays a meaningful role in shaping the information environment, consistent with the view that complexity hinders investors’ ability to process information, thereby increasing trading frictions and uncertainty.

Keywords: firm complexity; information asymmetry; global financial crisis.

JEL classification : G12, G14, G34

1. Introduction

Modern corporations are structurally complex, making it increasingly difficult for investors and intermediaries to interpret and process firm information (Loughran & McDonald, 2024). While extensive research links transparency and disclosure practices to market frictions (see, i.e., Goldstein & Yang, 2019; Christensen, Hail & Leuz, 2021; Wang, Yu & Zhang, 2023; Davani & Davani, 2024), direct evidence on how firm complexity itself shapes the information environment remains relatively limited. Understanding this relationship is economically important because information asymmetry underpins adverse selection, liquidity, and trading costs (Easley & O'Hara, 2004; Verrecchia, 2001). This study investigates the extent to which firm complexity drives information asymmetry.

From a theoretical perspective, market microstructure theory suggests that when investors struggle to interpret firm specific information, information asymmetry rises. Market makers protect themselves against adverse selection, and investors demand higher returns (Easley & O'Hara, 2004; Glosten & Milgrom, 1985). Firm complexity amplifies these frictions. It can arise from multi segment operations, extensive intangible assets, or lengthy and technical disclosures, all of which reduce transparency and hinder effective external monitoring (Bushman, Piotroski & Smith, 2004; Kim, Kraft & Ryan, 2013). These insights underscore that firm complexity is a defining characteristic of the contemporary information environment.

Empirical evidence further supports this view. Operational complexity, such as diversification across industries and geographies, complicates performance evaluation and increases coordination costs (Baik, Johnson, Kim & Yu, 2023; Cohen & Lou, 2012). Firms with multiple segments are particularly prone to selective disclosure and delayed price adjustment, consistent with greater valuation uncertainty (Binz, Hills & Kubic, 2023). On the financial reporting side, disclosure complexity, manifested as excessive length, low readability, and linguistic ambiguity, imposes additional barriers for investors. Such complexity is linked to higher costs of capital, greater earnings surprises, and reduced liquidity (Bonsall, Leone, Miller & Rennekamp, 2017; Ertugrul, Qiu & Wan, 2017; Li, 2008).

Furthermore, recent studies provide evidence that complexity affects market outcomes. Bae, Belo, Li, Lin and Zhao (2023) demonstrate that disclosure complexity increases volatility and reduces pricing efficiency. Clark, Palepu and Siddique (2024) show that machine learning based complexity scores predict wider bid ask spreads and higher probabilities of informed trading. Similarly, Loughran and McDonald (2024) show that their 10-K textual complexity measure is positively associated with higher audit fees, larger absolute unexpected earnings, and higher post-filing stock return volatility. Taken together, these findings indicate that both

operational and disclosure complexity elevate firm specific information risk, resulting in greater trading frictions. However, because complexity is inherently multidimensional, the absence of a comprehensive measure makes it challenging to fully capture its effects.

Traditional proxies capture different facets of complexity rather than the whole construct. For example, segment diversification captures processing difficulty for valuing multi-segment firms (Cohen & Lou, 2012); disclosure readability captures how hard the narrative is to parse, which managers can raise around benchmark-meeting years (Lo, Ramos & Rogo, 2017). Recent advances leverage textual analysis to capture nuanced complexity in disclosures (Hoberg & Phillips, 2016; Loughran & McDonald, 2016, 2024). Building on these advances, we construct a textual measure of firm complexity using earnings conference calls (ECCs), which offer real-time, interactive communication between managers and investors (Brown, Hillegeist & Lo, 2004; Price, Doran, Peterson & Bliss, 2012). ECCs are increasingly central to financial markets and represent a natural setting to measure linguistic and structural complexity, as they convey timely firm information and shape investor and analyst responses (Borochin, Cicon, DeLisle & Price, 2018).

Using 32,930 U.S. firm-year observations from 2002 to 2023, we find that firm complexity is positively associated with information asymmetry. The results remain robust after controlling for readability and employing alternative complexity proxies. The results also hold when using a residual measure of firm complexity, accounting for changes in complexity over time, and applying system GMM estimation. To strengthen identification, we exploit the Global Financial Crisis (GFC) as an exogenous shift in the complexity environment and document a mitigated complexity–asymmetry relation in the post-GFC period. A plausible mechanism is that, following the GFC, increased investor scrutiny, improved data infrastructure, and stronger regulatory oversight prompted firms to simplify their operations, reducing underlying complexity and, in turn, information asymmetry.

This paper contributes to research on firm complexity by offering large-sample, market-level evidence that complexity is directly associated with information asymmetry. While prior work has emphasized how disclosure quantity, tone, and readability affect investor processing information (Asay, Elliott & Rennekamp, 2017; Li, 2008; Tetlock, 2007), we focus on the structural attributes embedded within firms' disclosure frameworks. Our findings indicate that greater firm complexity captured by ECCs translates into higher information frictions. Our evidence complements the disclosure and transparency literature (Healy & Palepu, 2001; Verrecchia, 2001; Diamond & Verrecchia, 1991; Ferrer, Santamaría & Suarez, 2022) by

showing that, even when firms provide public information, the underlying firm complexity still materially influences how that information is processed and priced.

Second, we extend the market microstructure literature by showing that asymmetry is influenced not only by disclosure practices (Diamond & Verrecchia, 1991; Leuz & Verrecchia, 2000; Healy, Hutton & Palepu, 1999; Brown & Hillegeist, 2007) and governance arrangements (Chung, Elder & Kim, 2010; Prommin, Jumreornvong & Jiraporn, 2014), but also by the underlying level of complexity. We show that firm complexity imposes higher cognitive and processing costs on investors and intermediaries, which is reflected in greater information asymmetry. These are the classic manifestations of adverse selection and liquidity frictions. Importantly, we also show that after the GFC, firm complexity has reduced significantly and accordingly, the impact of firm complexity on information asymmetry has been attenuated.

Finally, we contribute methodologically by refining the Loughran–McDonald (LM) complexity measure with a name-aware filter. Several LM dictionary terms, e.g., *International / Intl*, *Worldwide*, *Global*, *Sovereign*, *Affiliate*, *Covenant*, *Venture*, *Security*, often appear in official firm names rather than signalling genuine operational complexity. This issue is especially salient in question and answer section (Q&A) exchanges, where analysts and managers frequently repeat the firm’s name.¹ To avoid mechanically inflating scores, we exclude these terms only when they occur as part of a firm’s legal or commonly used name and count them normally elsewhere. This targeted adjustment preserves construct validity while maintaining comparability with (i) LM-based studies, because we do not alter the LM dictionary or its tokenization, and (ii) the ECC literature that measures text from ECCs (e.g., Brown et al., 2004; Brockman, Li & Price, 2015). It also enhances replicability: researchers can implement the same transparent, firm-name filter without altering the underlying LM dictionary, ensuring that measured complexity reflects real processing difficulty rather than naming conventions.

¹ For example, firms whose legal names contain *International* (e.g., American International Group), *Worldwide* (e.g., C.H. Robinson Worldwide), *Global* (e.g., Global Payments), *Affiliated/Affiliate* (e.g., Affiliated Managers Group), *Covenant* (e.g., Covenant Logistics), *Venture* (e.g., Venture Global LNG), or *Security* (e.g., NAPCO Security Technologies; United Security Bancshares) are repeatedly referenced by analysts and management in Q&A turns. Illustratively, in Covenant Logistics’ Q4 2024 ECC Q&A, an executive stated, “*I think we’re going to be operating from a position of offense, not defence, be it whatever of those capital allocation alternatives makes the most sense for the long-term value creation for Covenant shareholders.*” A similar pattern appears in calls such as Orbitz Worldwide’s Q2 2014 ECC, where analysts and management repeatedly refer to “*Orbitz Worldwide*,” causing the token *Worldwide* to be flagged even though it is simply part of the firm’s name. To avoid mechanically inflating complexity, we exclude these tokens only when they occur as part of the issuer’s legal or commonly used name in the Q&A speaker text; occurrences elsewhere are counted normally.

The remainder of the paper is organized as follows. Section 2 describes the data, sample, and measures. Section 3 presents the main empirical results, robustness checks, endogeneity analyses, and GFC-period heterogeneity. Section 4 concludes.

2. Sample and data

2.1. Sample construction

The data for this study are drawn from two main sources. ECC transcripts from Refinitiv StreetEvents (Eikon) are used to construct textual measures of firm complexity, while information asymmetry measures, and firm-level accounting data come from the CRSP–Compustat merged database. Combining these sources yields a panel of 32,930 firm-year observations for U.S. listed firms from 2002 to 2023.

2.2. Measure of firm complexity

We assess firm complexity using ECC transcripts, which capture real-time communication between managers and investors and serve as a key information source for market participants. To quantify complexity, we apply the LM complexity dictionary to different sections of the transcripts,² allowing us to capture a comprehensive level of firm complexity.

First, we compute the Presentation Complexity Score (PCS) from the prepared remarks. This opening segment is scripted, written in advance, and coordinated with the earnings release and investor materials. Content typically includes an overview of the quarter, segment highlights, performance drivers (demand, costs, regulation), capital allocation decisions, and forward-looking guidance. Prepared remarks contain longer, more structured sentences and a consistent register around strategy and outlook. A higher PCS reflects the extent to which management conveys complex operational or financial details, such as multi-segment models, derivative usage, cross-border supply chains, or complex regulatory regimes, within an otherwise polished narrative. Because this portion is management-controlled, PCS captures intended firm complexity as deliberately communicated by management.

Second, we compute the Question and Answer Complexity Score (QCS) from the analyst–management exchange. This section is interactive and unscripted, allowing analysts to

² We note that the LM dictionary includes certain terms (e.g., *International*, *Global*, *Worldwide*) that frequently appear as part of firm names rather than as indicators of firm complexity. To avoid inflating scores, we exclude these occurrences only for firms where such terms are part of the official company name.

probe areas of uncertainty. Questions typically focus on assumptions behind guidance, unit economics, contract terms, regulatory exposures, acquisitions, liquidity, or hedging strategies. This section is shorter and more fragmented, with greater technical detail and topic switching. A higher QCS therefore reflects emergent firm complexity, the depth of operational or financial intricacy revealed only when management must respond spontaneously to investor questions.

Finally, we compute the Total Complexity Score (TCS) for the full transcript (prepared remarks and question and answer (Q&A) sections) to summarize overall complexity communicated across the entire event. This measure integrates both management's scripted narrative and spontaneous discussion, offering a comprehensive view of overall firm complexity. Specifically, definition of the section-level complexity score (S_c) is:

$$S_c = 100 \times \frac{1}{|W_s|} \sum_{w \in W_s} I(w \in C) \quad (1)$$

where $s \in \{\text{prepared remarks; Q\&A sections; and full transcript}\}$; W_s is the set of words in section s ; C is the LM complexity dictionary; and $I(w \in C) = 1$ if word w is in C , and 0 otherwise.

2.3. Measure of information asymmetry

We proxy for information asymmetry using two measures. First, we compute the bid–ask spread, a widely used indicator of information asymmetry that reflects market makers' compensation for adverse-selection risk (Glosten & Milgrom, 1985; Easley & O'Hara, 2004; Chung and Zhang, 2014). Following Chung and Zhang (2014), the bid–ask spread is defined as:

$$BIDASK_{it} = \frac{1}{D_{it}} \sum_{d=1}^{D_{it}} \left\{ \frac{Ask_{itd} + Bid_{itd}}{(Ask_{itd} + Bid_{itd})/2} \right\} \quad (2)$$

where D_{it} is the number of trading days for firm i in year t^3 , and Ask_{itd} and Bid_{itd} are the daily closing ask and bid quotes.

Our second measure of information asymmetry is return volatility. Firms with greater information asymmetry exhibit higher volatility in stock returns (Yang, Zhang & Zhang, 2020;

³ The number of trading days per calendar year is about 252 in U.S. markets (Zhang, Ruan & Zhang, 2023), but D_{it} varies by firm–year due to listings/delistings mid-year, trading halts, holidays, and missing quotes.

Chen, Huang & Jha, 2012). We calculate return volatility as the standard deviation (SD) of daily stock returns over the prior twelve months (Guo & Masulis, 2015).

2.4. Control variables

We include a comprehensive set of firm-level controls that are commonly used in the information asymmetry literature (e.g., Glosten & Milgrom, 1985; Easley & O'Hara, 2004). These include firm size (log of total assets), market-to-book assets (market value of assets relative to book value of assets), return on equity (net income scaled by common equity), leverage (total debt divided by assets), R&D intensity (R&D expenditure divided by sales, with missing values set to zero), dividend dummy (equal to one if a cash dividend is paid, zero otherwise), firm age, share price (log of period-end price), and trading volume (log of shares traded). In addition, we control for whether the firm is a “*regular caller*”, defined as hosting an ECC in the current year and at least one in any of the prior three years (Brown et al., 2004). Following standard practice, all continuous variables are winsorized at the 1st and 99th percentiles to mitigate the influence of outliers.

2.5. Summary statistics

Table 1 presents the descriptive statistics of the main variables. Return volatility averages 13.1%, with a median of 11.3% and a SD of 7.5%, which is consistent with Guo and Masulis (2015). The mean bid–ask spread is 4.8% (median 3.0%), which aligns with the levels reported in Chung and Zhang (2014). In our sample, 94.5% of firm-year observations correspond to regular callers, firms that host an ECC in the current period and at least once in any of the prior three years. This is consistent with the patterns documented by Brown et al. (2004).

Approximately 39.9% of firms pay dividends, and the average firm size (log assets = 6.897) is in line with Pan, Wang and Weisbach (2016). The R&D-to-sales ratio has a mean of 0.481 and a median of 0.007, reflecting the zero-inflated distribution of innovation expenditures noted by Azar, Kagy and Schmalz (2016). Average trading volume (log shares = 18.496) is comparable with previous findings (Brennan, Chordia & Subrahmanyam, 1998). The market-to-book ratio (assets) has a mean of 2.264, firm age averages log(age) = 2.513 (median 2.639), and share prices center around log(price) = 3.032; all of these measures are broadly consistent with distributions reported for U.S. listed firms (Bergers and Blomkvist, 2024).

[Please insert Table 1 here]

Table 2 presents pairwise correlations, showing that firm complexity is positively correlated with information asymmetry. While the correlations offer initial evidence consistent with our expectations, regression analysis is required to more rigorously estimate the effects of interest and account for potential confounding influences.

[Please insert Table 2 here]

3. Main empirical results

3.1. Basic results

We use the following regression model in our empirical analysis to examine the effect of firm complexity on information asymmetry:

$$\text{InfoAsym}_{it} = \alpha + \beta \text{Firm complexity}_{it-1} + \sum \text{Controls}_{it-1} + d_i + d_t + e_i \quad (3)$$

To address potential endogeneity concerns, we employ firm fixed effects (d_i) to control for unobservable, time-invariant firm characteristics. The inclusion of firm fixed effects enables us to identify the relationship between firm complexity and information asymmetry while controlling for within-firm variation over time. We further include year fixed effects (d_t) to capture macroeconomic shocks and market-wide trends. Additionally, we conduct lead-lag regressions to mitigate concerns regarding reverse causality. We cluster standard errors at firm level. In Eq. (3), the coefficient of interest is β , which measures the extent to which firm complexity measure from ECCs predicts firm's information asymmetry.

Table 3 presents the firm-fixed effect regression results for firm's information asymmetry on lagged firm complexity. Columns (1) and (2) reports results using PCS, Columns (3) and (4) reports results using QCS, and columns (5) and (6) reports results using TCS. Across all specifications, firm complexity exhibits positive and statistically significant coefficients, indicating that higher firm complexity is associated with greater information asymmetry. The results are also economically meaningful. For example, in Panel A, a one SD increase in firm complexity corresponds to about 1.6% of the mean return volatility and about 16.6% of the mean relative bid–ask spread.⁴ The control variables yield results consistent with prior literature. Information asymmetry decreases with firm size and dividend payment, consistent with larger and dividend-paying firms being more transparent (Pastor & Veronesi, 2003). In

⁴ Calculated as the estimated coefficient multiplied by one SD of the complexity measure, divided by the mean of the dependent variable (e.g., $0.0079 \times 0.27 \div 0.131 = 0.016$, or 1.6 % of the mean).

contrast, firms with higher R&D intensity and higher market-to-book ratios exhibit higher spreads and volatility, reflecting greater uncertainty about their future growth opportunities and the associated increased information asymmetry (Chan, Lakonishok & Sougiannis, 2001; Eberhart, Maxwell & Siddique, 2004; Pastor & Veronesi, 2003). Consistent with microstructure theory, trading volume and share prices are systematically related to liquidity: higher trading activity is associated with narrower spreads, and higher-priced stocks exhibit lower percentage spreads (Chordia, Roll & Subrahmanyam, 2001; Amihud & Mendelson, 1986; Stoll, 2000). Overall, the results in Table 3 are highly consistent with our expectations that firm complexity is positively associated with information asymmetry.

[Please insert Table 3 here]

3.2. Robustness checks

3.2.1. Addressing generic text readability

We control for the general readability of texts because readability both reflects firms' operational complexity and affects their information environment. Firms with complex operations tend to produce harder-to-read disclosures, with longer documents, longer sentences, and more complex vocabulary (Hwang & Kim, 2017). To capture these effects, we employ three different measures of text readability. We control for the Fog index, a widely used readability measure in financial text analysis (Bonsall et al., 2017). It accounts for readability while isolating firm complexity and avoiding omitted-variable bias. Longer sentences and a higher proportion of multi-syllable words make texts harder to read, and harder-to-read calls can independently increase trading frictions (e.g., by raising processing costs), even if the firm itself is not more complex. By including the Fog index, we ensure that our analysis captures the effect of firm complexity itself rather than the readability of the disclosure. Section-specific Fog indices are computed for the prepared remarks, the Q&A, and the full transcript. As shown in Table 4, after including these section-specific Fog indices as controls, the coefficients on PCS, QCS, and TCS remain positive and statistically significant in both the volatility regressions (Columns 1, 3, and 5) and the bid-ask spread regressions (Columns 2, 4, and 6). The effect sizes remain comparable to the baseline estimates: a one SD increase in complexity is associated with roughly 1–2% of the mean return volatility (PCS, QCS, and TCS) and about 10–18% of the mean relative spread. The findings indicate that the economic effects are both meaningful and consistently signed, even after accounting for readability. Overall, Table 4 confirms that our main findings reflect firm complexity rather than complexity attributable to text readability.

[Please insert Table 4 here]

Our second and third readability proxies are ECC transcript length and file size. Transcript length is measured as the total word count of the full ECC transcript (prepared remarks and Q&A sections), following the common practice of using document word counts as a complexity proxy in disclosure research (Li, 2008; You & Zhang, 2009). File size is measured as the size of the transcript file at retrieval (in kilobytes/megabytes), in the spirit of Loughran and McDonald (2014), who show that document file size is a simple, replicable readability proxy that performs well relative to traditional indices.

Each readability proxy is computed consistently within the relevant analysis; for example, when firm complexity is assessed for the prepared remarks, the corresponding readability measures are calculated for that section as well. Panels A and B of Table 5 show that, after controlling for transcript length and file size, the estimated effects of PCS, QCS, and TCS remain largely unchanged when the dependent variable is volatility (Columns 1, 3, and 5) or the bid ask spread (Columns 2, 4, and 6).

[Please insert Table 5 here]

3.2.2. Alternative measures of firm complexity

Beyond our section level dictionary measures, the literature offers several complementary proxies for firm complexity. To ensure that our findings are not driven by a single metric, we re-estimate the baseline models using three well established alternatives. First, we adopt the Loughran and McDonald Complexity Score (LMCS) as an external benchmark text-based measure (Loughran & McDonald, 2024). This dictionary is designed to capture processing difficulty in disclosures and provides an independent reference point to our section specific measures.⁵ Second, we use intangible intensity, defined as intangibles scaled by total assets, as a structural proxy for organizational opacity. Firms with a larger share of intangible assets are generally harder to value and monitor because these resources are non-physical, firm specific, and difficult to verify. Prior research links higher intangible intensity to increased information asymmetry and valuation frictions (Lev & Zarowin, 1999; Aboody & Lev, 2000; Barth, Kasznik & McNichols, 2001). Third, we construct a composite firm complexity index using a one factor model based on three standardized inputs: the number of business segments, firm size (log sales), and leverage. Each input is z scored, and the resulting factor score serves as a stable, aggregated proxy for firm complexity. This approach reduces

⁵ We also include LM file size (MB) to ensure effects are not driven by transcript data volume.

idiosyncratic measurement error and follows factor-based methods commonly applied in corporate finance (Coles, Daniel & Naveen, 2008).

For all three alternative measures, we maintain the same controls, fixed effects, and outlier treatment as in the baseline models (winsorization at the 1st and 99th percentiles by year). Consistent results across these proxies would indicate that the observed relationship between firm complexity and information frictions does not depend on a particular textual metric but instead reflects a broader underlying phenomenon.

We present these results in Table 6. Columns (1) and (2) show that LMCS is positively associated with both return volatility and bid–ask spread.⁶ Columns (3) and (4) indicate that firms with higher intangible asset intensity exhibit greater information asymmetry. In Columns (5) and (6), where firm complexity is captured by the composite firm complexity index, the results likewise show higher volatility and wider spreads. The magnitude and significance levels of all three firm complexity measures are broadly consistent with the baseline results, confirming that our findings are not driven by any single complexity construct. For example, when firm complexity is measured using intangibles, a one SD increase in intangible intensity corresponds to about 0.8% of the mean return volatility and about 38% of the mean relative bid–ask spread.⁷

[Please insert Table 6 here]

3.3. Endogeneity tests

To strengthen identification and address potential endogeneity, we apply four complementary checks. First, to mitigate the concern that our complexity measures might be driven by firms' fundamental attributes, we construct a residual firm-complexity measure by regressing each complexity score on firm characteristics and past financial performance and using the unexplained component as the key regressor (Borochin et al., 2018; Huang, Teoh & Zhang, 2014). Second, we analyze within-firm variation by estimating first-difference (change-on-change) specifications, which difference out time-invariant firm heterogeneity; the complexity–asymmetry relation remains positive and significant. Third, we estimate system-GMM models to allow for persistence in information asymmetry and potential reverse causality

⁶ LMCS coverage is available only through 2021, and this reduces the observations into 28,464 firm-years.

⁷ Calculated as the estimated coefficient multiplied by the SD of intangible intensity (0.2086), divided by the mean of the dependent variable, for example, $0.0053 \times 0.2086 \div 0.131 \approx 0.0084$ ($\approx 0.8\%$ of the mean volatility) and $0.0873 \times 0.2086 \div 0.048 \approx 0.379$ ($\approx 38\%$ of the mean spread). The mean of intangible intensity is 0.2015.

from information asymmetry to firm complexity. Finally, we exploit the Global Financial Crisis as a period of heightened investor scrutiny, regulatory oversight, and pressure for transparency and assess whether the complexity–asymmetry relationship changes in a tighter oversight environment. We find that firm complexity declines post-GFC and that the complexity–asymmetry relation weakens in the post-crisis period, consistent with firms simplifying in a tighter oversight environment.

3.3.1. Information asymmetry and residual firm complexity

To ensure that firm complexity measurement is not captured by firms' other fundamental attributes or pass financial performance, we isolate the component of firm complexity that is orthogonal to those factors and then test whether it still predicts information asymmetry. This approach lets us focus on the content-driven (unexpected) portion of the signal rather than variation explained by fundamentals or shared time effects.

For each fiscal year t , we residualize complexity scores in annual cross-sectional regressions. Specifically, for each fiscal year t :

$$S_{it}^c = \alpha_t + \beta'_t X_{it} + u_{it}^c, \quad c \in \{PCS, QCS, TCS\} \quad \text{define the residualized measures as}$$

$$\text{Re_PCS}_{it} = \hat{\varepsilon}_{it}, \text{Re_QCS}_{it} = \hat{v}_{it}, \text{and Re_TCS}_{it} = \hat{\eta}_{it} \quad (4)$$

where the dependent variable S_{it}^c is, in turn, each section-level complexity score (PCS, QCS, TCS) for firm i . The control vector X_{it} includes log total assets (size), market-to-book assets, return on equity, R&D-to-sales, dividend dummy, firm age, log share price, log trading volume, and the regular-caller indicator.

We then re-estimate the outcome model using the residualized measures:

$$\text{InfoAsym}_{it} = \theta \text{Re_Complexity}_{it-1} + \lambda' X_{it-1} + \mu_i + \tau_t + \epsilon_{i,t} \quad (5)$$

where InfoAsym_{it} is bid–ask spread or return volatility; μ_i and τ_t denote firm and year fixed effects. We clustered standard errors at firm level.

Because Re_Complexity is orthogonal to observables and year effects, θ captures the content-driven (unexpected) component of firm complexity. In our data, the coefficients on Re_Complexity (Re_PCS , Re_QCS , Re_TCS) remain positive and statistically significant for volatility and spreads, indicating that semantic complexity, rather than correlated fundamentals

or macro year shocks drives the complexity-asymmetry relation. These results report in Table 7.

[Please insert Table 7 here]

3.3.2. Information asymmetry and firm complexity change

To absorb time-invariant unobservable firm characteristics, we employ a within-firm change (first difference) analysis. If changes in complexity co-move with changes in information asymmetry at the same firm, that provides a more demanding, within-entity confirmation of our baseline relation.

We estimate:

$$\text{InfoAsym}_{it} = \beta_1 \Delta \text{Complexity}_{it-1} + \Delta \text{Controls}_{it-1} + \Delta \varepsilon_{it} \quad (6)$$

where $\Delta \text{Complexity}_{it-1}$ is defined as the firm complexity score (PCS, QCS, TCS.) in the current year compared with the previous year. We estimate regressions with firm and year fixed effects using the same control set as in the baseline and cluster standard errors at the firm level. This design reduces concerns about persistent confounds and focuses on year-to-year movements (Angrist & Pischke, 2009).

As reported in Table 8, the results show that if firm becomes more complexity in the current year than in the previous year, the firm will experience higher volatility and bid-ask spread. These within-firm results corroborate the baseline, showing that when a given firm becomes more complex, its information asymmetry rises accordingly, even after differencing out firm-fixed, time-invariant factors.

[Please insert Table 8 here]

3.3.3. System GMM

To address potential reverse causality between firm complexity and information asymmetry, we estimate a dynamic panel model using the system GMM estimator. Specifically, we include the one-period lag of the dependent variable (information asymmetry) as an additional regressor. The endogenous regressor of interest is the complexity measure (PCS, QCS, or TCS). Following best practice, lagged levels and differences of complexity and the dependent variables are used as internal instruments, with the instrument set collapsed to avoid proliferation (Roodman, 2009).

Table 9 reports the results, showing that complexity remains positively associated with information asymmetry after accounting for dynamic endogeneity. Diagnostic tests indicate valid instruments (Hansen test p-values above conventional thresholds) and no evidence of second-order autocorrelation (insignificant AR (2) tests), supporting the robustness of the specification.

[Please insert Table 9 here]

3.3.4. GFC period heterogeneity

The GFC exposed the fragility of complex, model-heavy frameworks, prompting markets and regulators to favor simpler, more transparent structures and robust rules (e.g., leverage-ratio backstops and “*simple, transparent, comparable*” securitisations) (Haldane, 2012). The crisis generated extreme market turbulence and heightened information frictions, and the subsequent recovery period was marked by regulatory reforms, stricter disclosure standards, and enhanced market monitoring (Haldane, 2012). Under this intensified oversight and investor scrutiny, excessive organizational complexity became a disadvantage. Empirical evidence indicates that firms responded by streamlining structures and reducing unnecessary complexity to restore trust and improve transparency (Carmassi & Herring, 2016; Correa & Goldberg, 2022). In this sense, the GFC acts as an exogenous shock that encouraged firms to lower their complexity, thereby mitigating its negative impact on the information environment.

To examine the evolution of firm complexity over time, we plot the average annual scores of three firm complexity measures from 2002 to 2023. Figure 1 shows that, for all three measures, complexity peaked around the GFC and subsequently declined over the following decade. These patterns provide preliminary empirical support for our argument.

[Please insert Figure 1 here]

A natural concern regarding the observed decline in firm complexity after the GFC is that it may simply reflect shorter transcripts rather than a genuine reduction in complexity. To address this, we plot trends in both total word counts and complexity word counts across the three measures. Figure 2 shows that while the number of complexity words declines notably, particularly after 2009, total word counts remain broadly stable and even increase in some years. Therefore, the post GFC decline in complexity is not driven by shorter ECCs but rather by a reduced use of complex language.

[Please insert Figure 2 here]

Therefore, the GFC provides an exogenous source of variation in firm complexity, allowing us to identify the pure effect of firm complexity on information asymmetry. Following prior studies that benchmark crisis-related periods (e.g., Kahle & Stulz, 2013; Lins, Servaes &

Tamayo, 2017), we define *Post_GFC* as 2011 onwards. We leave 2010 as a transition year to allow for policy implementation and market adjustment after the GFC.

We use *Post_GFC* in the interaction $\text{Complexity}_{it-1} \times \text{Post_GFC}$ to test whether the complexity–asymmetry relation has been mitigated after the crisis. Specifically, we use specification (7) to examine the impact of financial crisis on the relation between firm complexity and information asymmetry:

$$\text{InfoAsym}_{it} = \alpha + \beta_1 \text{Firm complexity}_{it-1} + \beta_2 \text{Post_GFC} + \beta_3 (\text{Firm complexity}_{it-1} \times \text{Post_GFC}) + \sum \text{Controls}_{it-1} + d_i + e_i \quad (7)$$

where InfoAsym_{it} is return volatility or relative spread, Complexity_{it-1} is the lagged firm complexity measures (PCS,QCS,TCS), *Post_GFC* equals 1 in the post-GFC period and 0 otherwise, X_{it-1} are lagged controls, d_i is firm fixed effects. Standard errors are clustered at the firm level. In this setup, β_1 captures the *Pre_GFC* complexity effect, and a negative β_3 indicates the reduced impact of firm complexity on information asymmetry after the crisis.

Table 10 presents the regression results: firm complexity is measured as PCS in columns (1)-(2), QCS in columns (3)-(4), and TCS in columns (5)-(6), with odd columns using return volatility and even columns using the relative bid–ask spread as the dependent variable. The main variable of interest is the interaction between complexity and the *Post_GFC* dummy. Across specifications, the coefficients on the interaction terms are negative and statistically significant, indicating that the positive association between complexity and information asymmetry is weaker after 2011. The economic magnitude of these effects is meaningful. For instance, A one SD increase in PCS has a smaller effect in the post-GFC period is about 2% of the mean return volatility and about 5.9% of the mean relative bid–ask spread.⁸ These results suggest that, following the financial crisis, improvements in the market environment and monitoring capacity have encouraged firms to reduce their complexity levels, thereby improving the firm’s information environment and reducing information asymmetry.

[Please insert Table 10 here]

Overall, the findings support the view that greater firm complexity is associated with higher information asymmetry. The GFC serves as an exogenous shock that reshaped the

⁸ Calculated using the interaction coefficients and $\text{SD}(\text{PCS})=0.270$: for volatility, $-0.0097 \times 0.270 \div 0.131 \approx 0.020$ (2% of the mean); for the spread, $-0.0104 \times 0.270 \div 0.048 \approx 0.059$ (5.9% of the mean).

market environment and strengthened monitoring capacity, encouraging firms to simplify their operations and structures. As a result, the post-crisis reduction in complexity contributed to an improved information environment and reduced information asymmetry.

4. Conclusion

This study investigates how firm complexity shapes information asymmetry in U.S. listed firms from 2002 to 2023. Using ECC transcripts to construct textual measures of firm complexity, we find that more complex firms exhibit higher information asymmetry, as captured by return volatility and bid–ask spreads. These results remain after controlling for readability and alternative firm complexity proxies, indicating that our measures capture complexity beyond generic text difficulty.

We further validate the robustness of our findings by (i) using a residual complexity measure that strips out variation explained by firm fundamentals, (ii) analyzing within-firm changes in complexity and information asymmetry over time, and (iii) estimating system GMM models to address persistence and potential reverse causality. In addition, we exploit the GFC as period of heightened scrutiny and pressure for transparency and show that firms reduced their complexity and that the complexity–information asymmetry relation weakened in the post-GFC period. Overall, the evidence suggests that firm complexity plays a meaningful role in shaping the information environment, consistent with the view that complexity hinders investors’ ability to process information, thereby increasing trading frictions and uncertainty.

Funding statement

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

- Aboody, D., & Lev, B. (2000). Information asymmetry, R&D, and insider gains. *The Journal of Finance*, 55(6), 2747-2766. <https://doi-org.surrey.idm.oclc.org/10.1111/0022-1082.00305>
- Asay, H. S., Elliott, W. B., & Rennekamp, K. (2017). Disclosure readability and the sensitivity of investors' valuation judgments to outside information. *The Accounting Review*, 92(4), 1-25. <https://www.jstor.org/stable/26550672>
- Amihud, Y., & Mendelson, H. (1986). Asset pricing and the bid-ask spread. *Journal of Financial Economics*, 17(2), 223-249. [https://doi-org.surrey.idm.oclc.org/10.1016/0304-405X\(86\)90065-6](https://doi-org.surrey.idm.oclc.org/10.1016/0304-405X(86)90065-6)
- Angrist, J. D., & Pischke, J. S. (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.
- Azar, J. A., Kagy, J.-F., & Schmalz, M. C. (2016). Can changes in the cost of carry explain the dynamics of corporate “cash” holdings? *Review of Financial Studies*, 29(8), 2194–2240. <https://doi-org.surrey.idm.oclc.org/10.1093/rfs/hhw021>
- Barth, M. E., Kasznik, R., & McNichols, M. F. (2001). Analyst coverage and intangible assets. *Journal of Accounting Research*, 39(1), 1-34. <https://doi-org.surrey.idm.oclc.org/10.1111/1475-679X.00001>
- Bae, J. W., Belo, F., Li, J., Lin, X., & Zhao, X. (2023). The opposing effects of complexity and information content on uncertainty dynamics: evidence from 10-K filings. *Management Science*, 69(10), 6313–6332. <https://doi.org/10.1287/mnsc.2022.4615>
- Baik, B., Johnson, M. F., Kim, K., & Yu, K. (2023). Organization complexity, financial reporting complexity, and firms' information environment. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.4413814>

- Bergers, O., & Blomkvist, M. (2024). The alphabet and idiosyncratic volatility. *Journal of Financial Research*, 47(2), 197–228. <https://doi-org.surrey.idm.oclc.org/10.1111/jfir.12369>
- Binz, O., Hills, R., & Kubic, M. (2023). Did the FASB codification reduce the complexity of applying U.S. GAAP? *Journal of Accounting Research*, 61(5), 1479–1530. <https://doi.org/10.1111/1475-679X.12480>
- Bonsall, S. B., Leone, A. J., Miller, B. P., & Rennekamp, K. (2017). A plain English measure of financial reporting readability. *Journal of Accounting and Economics*, 63(2–3), 329–357. <https://doi.org/10.1016/j.jacceco.2017.03.002>
- Borochin, P., Cicon, J., DeLisle, J., & Price, S. M. (2018). The effects of conference call content on market perceptions of value uncertainty. *Journal of Financial Markets*, 40, 75–91. <https://doi-org.surrey.idm.oclc.org/10.1016/j.finmar.2017.12.003>
- Brennan, M. J., Chordia, T., & Subrahmanyam, A. (1998). Alternative factor specifications, security characteristics, and the cross-section of expected stock returns. *Journal of Financial Economics*, 49(3), 345–373. [https://doi-org.surrey.idm.oclc.org/10.1016/S0304-405X\(98\)00028-2](https://doi-org.surrey.idm.oclc.org/10.1016/S0304-405X(98)00028-2)
- Brockman, P., Li, X., & Price, S. M. (2015). Differences in conference call tones: Managers vs. analysts. *Financial Analysts Journal*, 71(4), 24–42. <https://doi-org.surrey.idm.oclc.org/10.2469/faj.v71.n4.1>
- Brown, S., Hillegeist, S. A., & Lo, K. (2004). Conference calls and information asymmetry. *Journal of Accounting and Economics*, 37(3), 343–366. <https://doi.org/10.1016/j.jacceco.2004.02.001>
- Brown, S., & Hillegeist, S. A. (2007). How disclosure quality affects the level of information asymmetry. *Review of Accounting Studies*, 12(2), 443–477. <https://doi-org.surrey.idm.oclc.org/10.1007/s11142-007-9032-5>

- Bushman, R. M., Piotroski, J. D., & Smith, A. J. (2004). What determines corporate transparency? *Journal of Accounting Research*, 42(2), 207–252. <https://doi.org/10.1111/j.1475-679X.2004.00136.x>
- Carmassi, J., & Herring, R. (2016). The corporate complexity of global systemically important banks. *Journal of Financial Services Research*, 49(2–3), 175–201. <https://doi.org/10.1007/s10693-016-0251-4>
- Chan, L. K. C., Lakonishok, J., & Sougiannis, T. (2001). The stock market valuation of R&D expenditures. *Journal of Finance*, 56(6), 2431–2456. <https://www.jstor.org/stable/2697829>
- Chen, C., Huang, A. G., & Jha, R. (2012). Idiosyncratic return volatility and the information quality underlying managerial discretion. *Journal of Financial and Quantitative Analysis*, 47(4), 873–899. <https://doi.org/10.1017/S002210901200018X>
- Christensen, H. B., Hail, L., & Leuz, C. (2021). Mandatory CSR and sustainability reporting: Economic analysis and literature review. *Review of Accounting Studies*, 26(3), 1176–1248.
- Chordia, T., Roll, R., & Subrahmanyam, A. (2001). Market liquidity and trading activity. *The Journal of Finance*, 56(2), 501–530. <https://doi.org/10.1111/0022-1082.00335>
- Chung, K. H., Elder, J., & Kim, J. C. (2010). Corporate governance and liquidity. *Journal of Financial and Quantitative Analysis*, 45(2), 265–291. <https://doi.org/10.1017/S0022109010000104>
- Chung, K. H., & Zhang, H. (2014). A simple approximation of intraday spreads using daily data. *Journal of Financial Markets*, 17, 94–120. <https://doi-org.surrey.idm.oclc.org/10.1016/j.finmar.2013.02.004>

- Clark, B. J., Palepu, S., & Siddique, A. R. (2024). Firm complexity and information asymmetry: evidence from ML-based complexity to measure information processing costs. SSRN Electronic Journal. <https://doi.org/10.2139/ssrn.4763575>
- Cohen, L., & Lou, D. (2012). Complicated firms. *Journal of Financial Economics*, 104(2), 383-400. <https://doi.org/10.1016/j.jfineco.2011.08.006>
- Coles, J. L., Daniel, N. D., & Naveen, L. (2008). Boards: Does one size fit all?. *Journal of Financial Economics*, 87(2), 329-356. <https://doi-org.surrey.idm.oclc.org/10.1016/j.jfineco.2006.08.008>
- Correa, R., & Goldberg, L. S. (2022). Bank complexity, governance, and risk. *Journal of Banking and Finance*, 134, 106013. <https://doi.org/10.1016/j.jbankfin.2020.106013>
- Davani, S., & Davani, A. (2024). The Impact of New Information Disclosure on Information Asymmetry and Liquidity. Available at SSRN 5013109.
- Diamond, D. W., & Verrecchia, R. E. (1991). Disclosure, liquidity, and the cost of capital. *The Journal of Finance*, 46(4), 1325-1359. <https://doi-org.surrey.idm.oclc.org/10.1111/j.1540-6261.1991.tb04620.x>
- Easley, D., & O'hara, M. (2004). Information and the cost of capital. *The Journal of Finance*, 59(4), 1553-1583. <https://doi-org.surrey.idm.oclc.org/10.1111/j.1540-6261.2004.00672.x>
- Eberhart, A. C., Maxwell, W. F., & Siddique, A. R. (2004). An examination of long-term abnormal stock returns and operating performance following R&D increases. *Journal of Finance*, 59(2), 623–650. <https://doi-org.surrey.idm.oclc.org/10.1111/j.1540-6261.2004.00644.x>
- Ertugrul, M., Lei, J., Qiu, J., & Wan, C. (2017). Annual report readability, Tone ambiguity, and the cost of borrowing. *Journal of Financial and Quantitative Analysis*, 52(2), 811–836. <https://doi.org/10.1017/S0022109017000187>

- Ferrer, E., Santamaría, R., & Suárez, N. (2022). Complexity is never simple: Intangible intensity and analyst accuracy. *BRQ Business Research Quarterly*, 25(2), 143-172. <https://doi-org.surrey.idm.oclc.org/10.1177/2340944420931871>
- Glosten, L. R., & Milgrom, P. R. (1985). Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. *Journal of Financial Economics*, 14(1), 71–100. [https://doi.org/10.1016/0304-405X\(85\)90044-3](https://doi.org/10.1016/0304-405X(85)90044-3)
- Goldstein, I., & Yang, L. (2019). Good disclosure, bad disclosure. *Journal of Financial Economics*, 131(1), 118-138. <https://doi.org/10.1016/j.jfineco.2018.08.004>
- Guo, L., & Masulis, R. W. (2015). Board structure and monitoring: New evidence from CEO turnovers. *The Review of Financial Studies*, 28(10), 2770-2811. <https://www.jstor.org/stable/24466888>
- Haldane, A. G. (2012). The dog and the frisbee. Jackson hole economic policy symposium. Bank of England/BIS. <https://www.bankofengland.co.uk/paper/2012/the-dog-and-the-frisbee/>
- Healy, P. M., & Palepu, K. G. (2001). Information asymmetry, corporate disclosure, and the capital markets: A review of the empirical disclosure literature. *Journal of Accounting and Economics*, 31(1–3), 405–440. [https://doi.org/10.1016/S0165-4101\(01\)00018-0](https://doi.org/10.1016/S0165-4101(01)00018-0)
- Healy, P. M., Hutton, A. P., & Palepu, K. G. (1999). Stock performance and intermediation changes surrounding sustained increases in disclosure. *Contemporary Accounting Research*, 16(3), 485-520. <https://doi-org.surrey.idm.oclc.org/10.1111/j.1911-3846.1999.tb00592.x>
- Hoberg, G., & Phillips, G. (2016). Text-based network industries and endogenous product differentiation. *Journal of Political Economy*, 124(5), 1423-1465. <https://doi.org/10.1086/688176>

- Huang, X., Teoh, S. H., & Zhang, Y. (2014). Tone management. *The Accounting Review*, 89(3), 1083–1113. <https://www.jstor.org/stable/24468328>
- Hwang, B. H., & Kim, H. H. (2017). It pays to write well. *Journal of Financial Economics*, 124(2), 373-394. <http://dx.doi.org/10.1016/j.jfineco.2017.01.006>
- Kahle, K. M., & Stulz, R. M. (2013). Access to capital, investment, and the financial crisis. *Journal of Financial Economics*, 110(2), 280-299. <https://doi-org.surrey.idm.oclc.org/10.1016/j.jfineco.2013.02.014>
- Kim, S., Kraft, P., & Ryan, S. G. (2013). Financial statement comparability and credit risk. *Review of Accounting Studies*, 18(3), 783–823. <https://doi.org/10.1007/s11142-013-9233-z>
- Leuz, C., & Verrecchia, R. E. (2000). The economic consequences of increased disclosure. *Journal of Accounting Research*, 91-124. <https://www.jstor.org/stable/2672910>
- Lev, B., & Zarowin, P. (1999). The boundaries of financial reporting and how to extend them. *Journal of Accounting Research*, 37(2), 353-385. <https://doi.org/10.2307/2491413>
- Li, F. (2008). Annual report readability, current earnings, and earnings persistence. *Journal of Accounting and Economics*, 45(2–3), 221–247. <https://doi.org/10.1016/j.jacceco.2008.02.003>
- Lins, K. V., Servaes, H., & Tamayo, A. (2017). Social capital, trust, and firm performance: The value of corporate social responsibility during the financial crisis. *the Journal of Finance*, 72(4), 1785-1824. <https://doi-org.surrey.idm.oclc.org/10.1111/jofi.12505>
- Lo, K., Ramos, F., & Rogo, R. (2017). Earnings management and annual report readability. *Journal of Accounting and Economics*, 63(1), 1–25. <https://doi.org/10.1016/j.jacceco.2016.09.002>

- Loughran, T., & McDonald, B. (2014). Measuring readability in financial disclosures. *the Journal of Finance*, 69(4), 1643-1671. <https://www.jstor.org/stable/43611199>
- 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>
- Loughran, T., & McDonald, B. (2024). Measuring firm complexity. *Journal of Financial and Quantitative Analysis*, 59(6), 2487–2514. <https://doi.org/10.1017/S0022109023000716>
- Pan, Y., Wang, T. Y., & Weisbach, M. S. (2016). CEO investment cycles. *Review of Financial Studies*, 28(10), 2770-2811. <https://www.jstor.org/stable/24466888>
- Pastor, L., & Veronesi, P. (2003). Stock valuation and learning about profitability. *Journal of Finance*, 58(5), 1749–1790. <https://doi-org.surrey.idm.oclc.org/10.1111/1540-6261.00587>
- Price, S. M., Doran, J. S., Peterson, D. R., & Bliss, B. A. (2012). Earnings conference calls and stock returns: The incremental informativeness of textual tone. *Journal of Banking & Finance*, 36(4), 992-1011. <https://doi.org/10.1016/j.jbankfin.2011.10.013>
- Prommin, P., Jumreornvong, S., & Jiraporn, P. (2014). The effect of corporate governance on stock liquidity: The case of Thailand. *International Review of Economics & Finance*, 32, 132-142. <https://doi-org.surrey.idm.oclc.org/10.1016/j.iref.2014.01.011>
- Roodman, D. (2009). How to do xtabond2: An introduction to difference and system GMM in Stata. *The Stata Journal*, 9(1), 86–136. <https://doi-org.surrey.idm.oclc.org/10.1177/1536867X0900900106>
- Stoll, H. R. (2000). Presidential address: friction. *The Journal of Finance*, 55(4), 1479-1514. 002 <https://doi-org.surrey.idm.oclc.org/10.1111/0022-1082.00259>

- Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. *The Journal of finance*, 62(3), 1139-1168. <https://doi-org.surrey.idm.oclc.org/10.1111/j.1540-6261.2007.01232.x>
- Verrecchia, R. E. (2001). Essays on disclosure. *Journal of Accounting and Economics*, 32(1–3), 97–180. [https://doi-org.surrey.idm.oclc.org/10.1016/S0165-4101\(01\)00025-8](https://doi-org.surrey.idm.oclc.org/10.1016/S0165-4101(01)00025-8)
- Wang, K., Yu, X., & Zhang, B. (2023). Panda games: corporate disclosure in the eclipse of search. *Management Science*, 69(6), 3263-3284. <https://doi-org.surrey.idm.oclc.org/10.1287/mnsc.2022.4521>
- Yang, Y. C., Zhang, B., & Zhang, C. (2020). Is information risk priced? Evidence from abnormal idiosyncratic volatility. *Journal of Financial Economics*, 135(2), 528-554. <https://doi.org/10.1016/j.jfineco.2019.06.013>
- You, H., & Zhang, X. J. (2009). Financial reporting complexity and investor underreaction to 10-K information. *Review of Accounting Studies*, 14(4), 559-586. <http://link.springer.com/10.1007/s11142-008-9083-2>
- Zhang, J., Ruan, X., & Zhang, J. E. (2023). Do short-term market swings improve realized volatility forecasts?. *Finance Research Letters*, 58, 104629. <https://doi-org.surrey.idm.oclc.org/10.1016/j.frl.2023.104629>

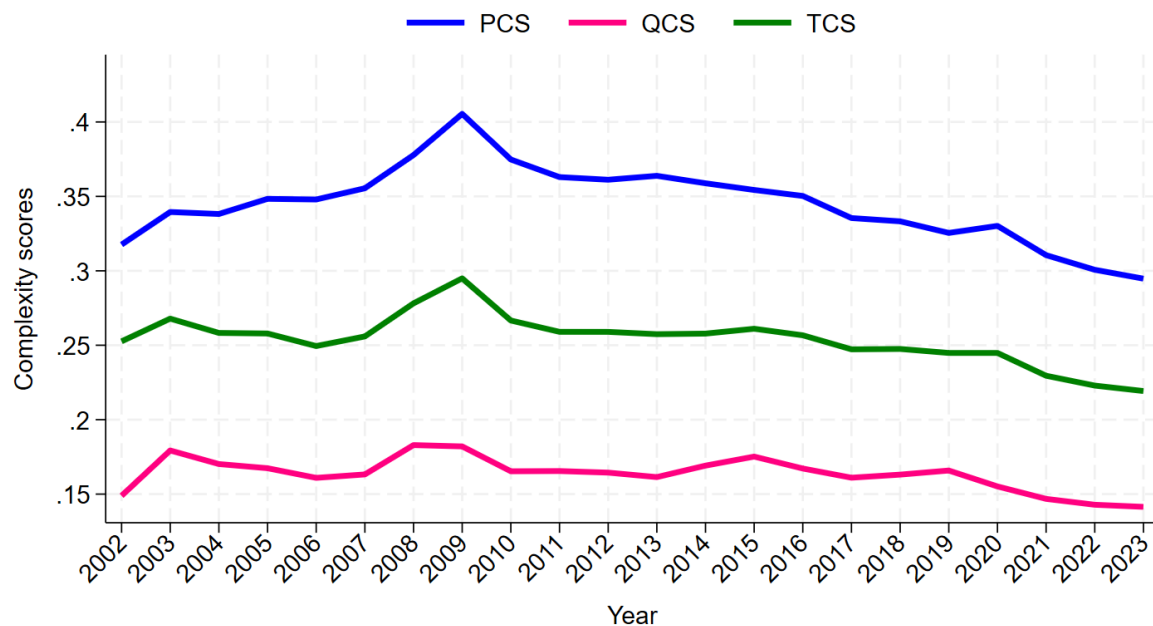
Appendix A. Variable definitions.

Variable	Variable description	Source
<i>Dependent variables</i>		
Return volatility	Standard deviation of monthly stock returns over the prior twelve months. Constructed from CRSP monthly returns.	CRSP/Compustat merged database
Bid–ask spread	Average quoted spread relative to the mid-point price, measured using the high (<i>ASKHI</i>) and low (<i>BIDLO</i>) quotes from CRSP.	
<i>Independent variables</i>		
PCS (percent)	Percentage of words flagged as complex in the prepared remarks (presentation) section of ECCs, using the Loughran and McDonald (2024) dictionary.	ECCs from Refinitiv Eikon
QCS (percent)	Percentage of complex words in the management Q&A section of ECCs.	
TCS (percent)	Percentage of complex words in the entire ECC transcript, including presentation and Q&A sections.	
<i>Firm characteristics variables</i>		
Regular caller	Dummy variable equal to one if the firm hosts a call in the current period and held at least one call in any of the prior three periods (Brown et al., 2004), and zero otherwise.	CRSP/Compustat merged database
Dividend	Dummy equal to 1 if cash dividends (<i>DIV</i>) are positive, 0 otherwise.	
Research & development	R&D expenditure (<i>XRD</i>) scaled by sales; missing values set to 0.	
Trading volume (log)	The natural logarithm annual shares traded (<i>CSHTR_F</i> + 1).	
Share price (log)	The natural logarithm of the fiscal year-end stock price (<i>PRCC_F</i>).	
Firm size (log)	The natural logarithm of total assets (<i>AT</i>).	
Return on equity	Net income divided by common equity.	
Market to book assets	The ratio of the market value of assets to the book value of assets. Market value of assets is measured as total assets minus book equity plus market value of equity.	
	$(AT - CEQ + (CSHO \times PRCCF)) / AT$	

Firm age (log)	The natural logarithm of the number of years since the firm first appears in the CRSP–Compustat linking table (<i>LINKDT</i>). Constructed as $\log(1 + (\text{year}(\text{DATADATE}) - \text{year}(\text{LINKDT})))$.	
<i>Other variables</i>		
Post-GFC	Dummy variable equal to 1 for firm-years 2011 and after, and 0 otherwise.	
Presentation Fog index	Gunning Fog readability index computed from the prepared (presentation) remarks section of the ECC transcript. Calculated as $0.4 \times (\text{average sentence length} + \text{percentage of complex words } [\geq 3 \text{ syllables}])$.	
Q&A Fog index	Gunning Fog readability index computed from the Q&A section of the ECC transcript of the interactive discussion.	
Total Fog index	Gunning Fog readability index computed from the full ECC transcript, combining the presentation (prepared remarks) and Q&A sections.	
Presentation transcript length	Word count of prepared remarks.	
Q&A transcript length	Word count of Q&A section.	
Total transcript length	Word count of full transcript.	
Presentation file size (MB)	File size of prepared-remarks section (megabytes).	
Q&A file size (MB)	File size of Q&A section (MB).	
Total file size (MB)	File size of Q&A section (MB).	
LMCS	Loughran–McDonald Complexity dictionary. https://sraf.nd.edu/loughranmcdonald-master-dictionary/	
Intangible intensity	Intangible assets / total assets (<i>INTAN/AT</i>).	CRSP/Compustat merged database

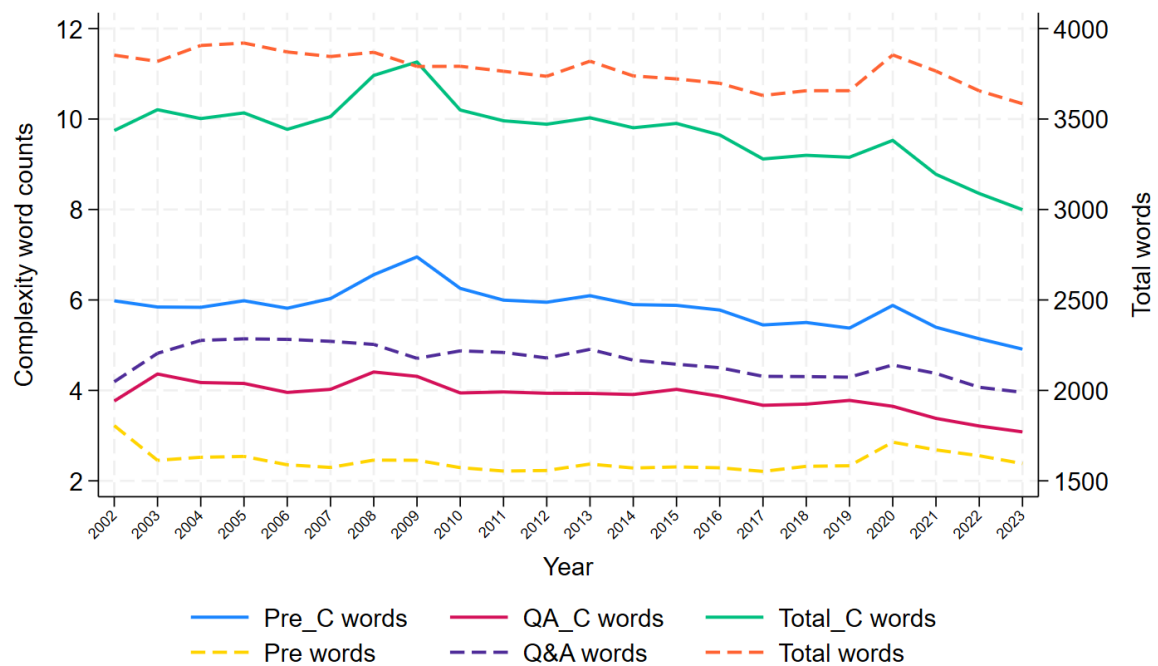
This table presents the definitions of all main variables used in our analysis. Mnemonics for Compustat variables are presented in parentheses, italicized and capitalized.

Figure 1. Trends in complexity scores, 2002–2023



This figure shows the average annual complexity scores from ECC transcripts, measured as the percentage of words flagged as complex using the LM dictionary. PCS (blue) represents complexity in prepared remarks, QCS (pink) represents complexity in the Q&A section, and TCS (green) represents overall transcript complexity. Prepared remarks consistently exhibit the highest complexity, while Q&A sections are the least complex. Complexity scores peak around 2009, coinciding with the GFC, and show a gradual decline in the subsequent decade.

Figure 2. Trends in complexity word counts and total words, 2002–2023



This figure plots average annual complexity word counts (solid lines, left axis) and total word counts (dashed lines, right axis) for ECC transcripts. “*Pre_C words*” and “*Pre words*” represent complexity words and total words in the prepared remarks section, “*QA_C words*” and “*Q&A words*” represent complexity words and total words in the Q&A section, and “*Total_C words*” and “*Total words*” represent overall transcript-level complexity and total words.

Table 1. Summary statistics

	Count	Min	P25	Mean	Median	P75	Max	SD
Return volatility	32,930	0.0313	0.0805	0.1311	0.1125	0.1605	0.6970	0.0750
Bid-ask spread	32,930	0.0009	0.0155	0.0477	0.0304	0.0549	0.7704	0.0645
PCS (percent)	32,930	0.0000	0.1600	0.3606	0.2979	0.4981	1.4690	0.2699
QCS (percent)	32,930	0.0000	0.0572	0.1639	0.1219	0.2221	1.1070	0.1551
TCS (percent)	32,930	0.0000	0.1207	0.2574	0.2129	0.3451	1.0953	0.1872
Regular caller	32,930	0.0000	1.0000	0.9455	1.0000	1.0000	1.0000	0.2271
Dividend	32,930	0.0000	0.0000	0.3985	0.0000	1.0000	1.0000	0.4896
Research & development	32,930	0.0000	0.0000	0.4813	0.0073	0.1016	88.9063	3.6995
Trading volume (log)	32,930	11.3626	17.4808	18.4955	18.4932	19.5635	24.3872	1.6027
Share price (log)	32,930	0.0588	2.2618	3.0323	3.1232	3.8118	8.1123	1.1333
Firm size (log)	32,930	1.0028	5.6084	6.8975	6.8718	8.1320	13.0448	1.8829
Return on equity	32,930	-10.8772	-0.0777	-0.0221	0.0787	0.1709	8.0521	0.9060
Market to book assets	32,930	0.5142	1.2275	2.2642	1.6773	2.5928	17.3956	1.7667
Firm age (log)	32,930	0.0000	1.7918	2.5130	2.6391	3.2581	4.1109	0.9945

The sample consists of 32,930 firm-year observations from 2002–2023. Accounting data are from Compustat; returns, prices, quotes, and volume from CRSP; and ECC transcript text (complexity scores) from Refinitiv StreetEvents (Eikon). Return volatility is the standard deviation of daily returns over the prior twelve months. Bid-ask spread is the average quoted spread relative to the midpoint price. PCS, QCS, and TCS are the percentages of words flagged as complex in the Presentation, Q&A, and Total transcript sections, respectively, using the LM dictionary. Regular caller equals one if the firm hosts a call in the current period and had at least one call in any of the prior three periods, and zero otherwise. Dividend equals one if cash dividends are positive in the period. Research and development expenditure scaled by sales. Trading volume is the natural log of shares traded. Share price is the natural log of the period-end price. Firm size is the natural log of total assets. Return on equity is net income relative to common equity. Market-to-book assets is market value of assets relative to book value of assets. Firm age is years since the firm’s first CRSP listing (reported in logs). All continuous variables are winsorized at the 1st and 99th percentiles (by year).

Table 2. Correlations

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Return volatility	1.00													
Bid-ask spread	0.09***	1.00												
PCS (percent)	0.07***	0.01*	1.00											
QCS (percent)	0.06***	0.01*	0.55***	1.00										
TCS (percent)	0.04***	0.01*	0.89***	0.80***	1.00									
Regular caller	-0.01*	0.03***	0.01	0.01	0.00	1.00								
Dividend	-0.34***	0.08***	0.07***	0.08***	0.07***	0.01	1.00							
Research & development	0.16***	-0.03***	-0.06***	-0.06***	-0.07***	0.01*	-0.10***	1.00						
Trading volume (log)	0.01	0.18***	0.09***	0.10***	0.07***	0.11***	0.13***	0.02***	1.00					
Share price (log)	-0.47***	0.65***	0.07***	0.06***	0.03***	0.03***	0.34***	-0.11***	0.24***	1.00				
Firm size (log)	-0.39***	0.34***	0.15***	0.15***	0.12***	0.07***	0.43***	-0.13***	0.67***	0.62***	1.00			
Return on equity	-0.17***	0.09***	0.04***	0.03***	0.03***	0.00	0.13***	-0.11***	0.03***	0.20***	0.17***	1.00		
Market to book assets	0.06***	0.36***	-0.05***	-0.04***	-0.06***	-0.04***	-0.13***	0.09***	0.12***	0.29***	-0.15***	-0.02***	1.00	
Firm age (log)	-0.27***	0.10***	0.12***	0.07***	0.09***	0.21***	0.34***	-0.10***	0.14***	0.25***	0.31***	0.11***	-0.16***	1.00

This table reports the pairwise correlations of the main variables. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Table 3. Baseline regressions

	(1)	(2)	(3)	(4)	(5)	(6)
PCS	0.0079*** (0.000)	0.0295* (0.093)				
QCS			0.0087*** (0.002)	0.0371* (0.091)		
TCS					0.0156*** (0.000)	0.0618** (0.025)
Regular caller	-0.0039* (0.068)	0.0950*** (0.000)	-0.0039* (0.069)	0.0950*** (0.000)	-0.0038* (0.076)	0.0954*** (0.000)
Dividend	-0.0032** (0.018)	-0.0429** (0.010)	-0.0032** (0.017)	-0.0431*** (0.010)	-0.0032** (0.016)	-0.0431*** (0.010)
Research & development	0.0003** (0.043)	0.0015** (0.046)	0.0003** (0.044)	0.0015** (0.047)	0.0003** (0.042)	0.0015** (0.045)
Trading volume	0.0074*** (0.000)	-0.0895*** (0.000)	0.0073*** (0.000)	-0.0897*** (0.000)	0.0074*** (0.000)	-0.0893*** (0.000)
Share price	-0.0311*** (0.000)	0.3227*** (0.000)	-0.0312*** (0.000)	0.3223*** (0.000)	-0.0309*** (0.000)	0.3235*** (0.000)
Firm size	-0.0011 (0.443)	0.1507*** (0.000)	-0.0009 (0.519)	0.1513*** (0.000)	-0.0011 (0.439)	0.1506*** (0.000)
Return on equity	-0.0021*** (0.000)	-0.0065* (0.081)	-0.0021*** (0.000)	-0.0064* (0.082)	-0.0021*** (0.000)	-0.0065* (0.081)
Market to book assets	0.0063*** (0.000)	0.0850*** (0.000)	0.0063*** (0.000)	0.0850*** (0.000)	0.0063*** (0.000)	0.0848*** (0.000)
Firm age	-0.0060*** (0.000)	0.0120 (0.293)	-0.0060*** (0.000)	0.0121 (0.290)	-0.0060*** (0.000)	0.0120 (0.293)
Firm fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	32,930	32,930	32,930	32,930	32,930	32,930
Adjusted R-squared	0.601	0.703	0.601	0.703	0.602	0.703

This table reports regression results of return volatility and the bid–ask spread on firm complexity measured from ECC transcripts. Return volatility is the dependent variable in Columns 1, 3, and 5, and the bid–ask spread is the dependent variable in Columns 2, 4, and 6. The key explanatory variables are the ECC-based complexity scores, PCS, QCS, and TCS, which are included separately across the corresponding column pairs. Controls include firm characteristics, and fixed effects for firm and year. Variables are winsorized at the 1st and 99th percentiles. Standard errors are clustered by firm. ***, **, * denote 1%, 5%, 10% significance.

Table 4. Robustness: Controlling for generic text readability (Fog index)

	(1)	(2)	(3)	(4)	(5)	(6)
PCS	0.0071*** (0.000)	0.0273* (0.086)				
QCS			0.0088*** (0.002)	0.0297* (0.090)		
TCS					0.0142*** (0.000)	0.0452* (0.092)
Presentation Fog index	0.0019*** (0.000)	0.0051* (0.080)				
Q&A Fog index			0.0013*** (0.002)	0.0153*** (0.000)		
Total Fog index					-0.0000 (0.871)	0.0035** (0.024)
Control variables as in Table 3	Yes	Yes	Yes	Yes	Yes	Yes
Firm fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	32,930	32,930	32,930	32,930	32,930	32,930
Adjusted R-squared	0.602	0.703	0.602	0.703	0.601	0.703

This table reports regressions of information asymmetry on transcript-based complexity scores (PCS, QCS, and TCS), with section-level Fog indices included as additional controls. The dependent variables are return volatility in Columns 1, 3, and 5, and the bid–ask spread in Columns 2, 4, and 6. All regressions include firm and year fixed effects, with standard controls. Continuous variables are winsorized at the 1st and 99th percentiles. Robust standard errors are clustered at the firm level. *, **, and *** denote significance at the 10%, 5%, and 1% levels.

Table 5. Robustness: Controlling for generic text readability (Transcript length and file size)

	Panel A					
	(1)	(2)	(3)	(4)	(5)	(6)
PCS	0.0082*** (0.000)	0.0299* (0.097)				
QCS			0.0108*** (0.000)	0.0407* (0.092)		
TCS					0.0151*** (0.000)	0.0596** (0.032)
Presentation transcript length (words)	-0.0010 (0.393)	-0.0019 (0.891)				
Q&A transcript length (words)			-0.0018*** (0.000)	-0.0031 (0.297)		
Total transcript length (words)					-0.0049*** (0.003)	-0.0201 (0.137)
Control variables as in Table 3	Yes	Yes	Yes	Yes	Yes	Yes
Firm fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	32,930	32,930	32,930	32,930	32,930	32,930
Adjusted R-squared	0.601	0.703	0.602	0.703	0.602	0.703

	Panel B					
	(1)	(2)	(3)	(4)	(5)	(6)
PCS	0.0080*** (0.000)	0.0284* (0.084)				
QCS			0.0091*** (0.001)	0.0394* (0.088)		
TCS					0.0153*** (0.000)	0.0601** (0.030)
Presentation file size (MB)	-0.0182 (0.788)	0.6205 (0.414)				
Q&A file size (MB)			-0.0962*** (0.004)	-0.4815* (0.091)		
Total file size (MB)					-0.0819** (0.025)	-0.4322 (0.203)
Control variables as in Table 3	Yes	Yes	Yes	Yes	Yes	Yes
Firm fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	32,930	32,930	32,930	32,930	32,930	32,930
Adjusted R-squared	0.601	0.703	0.601	0.703	0.602	0.703

This table reports regressions of information asymmetry on firm complexity measures (PCS, QCS, and TCS), adding section-level transcript length (word counts) and file size (MB) as additional controls. Panel A (Columns 1-6) includes

transcript length controls, while Panel B (Columns 1-6) includes file size controls. The dependent variables are return volatility in Columns 1, 3, and 5, and the bid–ask spread in Columns 2, 4, and 6. All models include standard controls, firm and year fixed effects. Variables are winsorized at the 1st and 99th percentiles. Robust standard errors are clustered at the firm level. *, **, *** denote significance at the 10%, 5%, and 1% levels.

Table 6. Robustness: Alternative measures of firm complexity (LMCS, intangibles, composite index)

	(1)	(2)	(3)	(4)	(5)	(6)
LMCS	0.0125*** (0.001)	0.0335* (0.092)				
Intangible			0.0053* (0.079)	0.0873* (0.089)		
Composite firm-complexity index					0.0074*** (0.006)	0.0545** (0.020)
LM file size	0.0036*** (0.004)	0.0012 (0.910)				
Control variables as in Table 3	Yes	Yes	Yes	Yes	Yes	Yes
Firm fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	28,464	28,464	32,930	32,930	32,506	32,506
Adjusted R-squared	0.599	0.707	0.601	0.703	0.599	0.705

This table reports regressions of information asymmetry on LMCS, intangibles, a composite firm-complexity index, and LM file size. The dependent variables are return volatility in Columns 1, 3, and 5, and the bid–ask spread in Columns 2, 4, and 6. All models include standard controls, firm and year fixed effects. Continuous variables are winsorized at the 1st and 99th percentiles. Robust standard errors are clustered at the firm level. *, **, *** denote significance at the 10%, 5%, and 1% levels.

Table 7. Residualized firm complexity scores

	(1)	(2)	(3)	(4)	(5)	(6)
Re_PCS	0.0076*** (0.000)	0.0384** (0.043)				
Re_QCS			0.0075*** (0.008)	0.0432* (0.098)		
Re_TCS					0.0144*** (0.000)	0.0697** (0.011)
Control variables as in Table 3	Yes	Yes	Yes	Yes	Yes	Yes
Firm fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	32,930	32,930	32,930	32,930	32,930	32,930
Adjusted R-squared	0.601	0.703	0.601	0.703	0.703	0.602

This table reports regressions of information asymmetry on residualized firm complexity measures (PCS, QCS, and TCS). The residuals are obtained from year-by-year cross-sectional regressions of each complexity measure on the baseline control variables. The dependent variables are return volatility in Columns 1, 3, and 5, and the bid–ask spread in Columns 2, 4, and 6. All specifications include the same controls, firm and year fixed effects, and firm-clustered robust standard errors. Continuous variables are winsorized at the 1st and 99th percentiles. *, **, *** denote 10%, 5%, 1% significance.

Table 8. Changes in firm complexity (within-firm estimation)

	(1)	(2)	(3)	(4)	(5)	(6)
Δ PCS	0.0030*	0.0018*				
	(0.094)	(0.086)				
Δ QCS			0.0055**	0.0044*		
			(0.044)	(0.082)		
Δ TCS					0.0077***	0.0041*
					(0.006)	(0.079)
Δ Control variables as in Table 3	Yes	Yes	Yes	Yes	Yes	Yes
Firm fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	28,780	28,780	28,780	28,780	28,780	28,780
Adjusted R-squared	0.227	0.189	0.227	0.189	0.227	0.189

This table reports regressions of information asymmetry on changes in complexity scores (Δ PCS, Δ QCS, and Δ TCS). The dependent variables are return volatility in Columns 1, 3, and 5, and the bid–ask spread in Columns 2, 4, and 6. All models include changes in standard controls, plus firm and year fixed effects. Continuous variables are winsorized at the 1st and 99th percentiles. Robust firm-clustered standard errors are in parentheses. *, **, *** denote significance at the 10%, 5%, and 1% levels.

Table 9. Dynamic panel estimation using System GMM

	(1)	(2)	(3)	(4)	(5)	(6)
PCS	0.0384*** (0.000)	0.0215* (0.097)				
QCS			0.0872*** (0.000)	0.1221* (0.100)		
TCS					0.0600*** (0.000)	0.0257* (0.092)
L. Return volatility	0.2520*** (0.000)		0.2147*** (0.000)		0.2507*** (0.000)	
L. Bid-ask Spread		0.6014*** (0.000)		0.6243*** (0.000)		0.6083*** (0.000)
Constant	0.1081*** (0.000)	-0.0424 (0.351)	0.1036*** (0.000)	-0.0442 (0.315)	0.1137*** (0.000)	-0.0495 (0.272)
AR (1) test (p-value)	0.000	0.000	0.000	0.000	0.000	0.000
AR (2) test (p-value)	0.215	0.215	0.215	0.215	0.215	0.215
Hansen over-identification test (p-value)	0.328	0.328	0.328	0.328	0.328	0.328
Control variables as in Table 3	Yes	Yes	Yes	Yes	Yes	Yes
Firm fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	28,781	28,781	28,781	28,781	28,781	28,781

This table reports two-step System GMM estimates of return volatility in Columns 1, 3, and 5, and the bid–ask spread in Columns 2, 4, and 6, with PCS, QCS, and TCS as the key explanatory variables. All specifications include the lagged dependent variable and standard control variables. Continuous variables are winsorized at the 1st and 99th percentiles. Robust two-step standard errors are in parentheses. *, **, *** denote significance at 10%, 5%, and 1%.

Table 10. Post-GFC heterogeneity

	(1)	(2)	(3)	(4)	(5)	(6)
PCS	0.0106*** (0.000)	0.0040* (0.087)				
PCS * <i>Post_GFC</i>	-0.0097*** (0.003)	-0.0104* (0.087)				
QCS			0.0129*** (0.007)	0.0367* (0.093)		
QCS * <i>Post_GFC</i>			-0.0077 (0.135)	-0.0444* (0.084)		
TCS					0.0219*** (0.000)	0.0393* (0.098)
TCS * <i>Post_GFC</i>					-0.0132*** (0.005)	-0.0116* (0.080)
Post-GFC	-0.0148*** (0.000)	-0.0678*** (0.000)	-0.0173*** (0.000)	-0.0641*** (0.000)	-0.0150*** (0.000)	-0.0688*** (0.000)
Control variables as in Table 3	Yes	Yes	Yes	Yes	Yes	Yes
Firm fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	32,930	32,930	32,930	32,930	32,930	32,930
Adjusted R-squared	0.520	0.671	0.519	0.671	0.520	0.671

This table reports regressions of information asymmetry on complexity scores (PCS, QCS, and TCS) and their interactions with a financial crisis dummy. The dependent variables are return volatility in Columns 1, 3, and 5, and the bid–ask spread in Columns 2, 4, and 6. All models include standard controls, firm and year fixed effects. Continuous variables are winsorized at the 1st and 99th percentiles. Robust standard errors are clustered at the firm level. *, **, *** denote significance at the 10%, 5%, and 1% levels.