WHAT THE MARKET READS

HOW EU DISCLOSURE HARMONIZATION RESHAPED THE PRICING OF SWEDISH IPO PROSPECTUSES

HEDDA GUMMESSON
GUSTAV JÄGER

Bachelor Thesis
Stockholm School of Economics
2026

What the Market Reads

Abstract:
This paper studies whether the EU Prospectus Regulation (EU) 2017/1129 changed the way Swedish IPO prospectuses are written, and whether these changes affect first-day underpricing. We analyze 269 Swedish IPOs between 2006 and 2026 and use the textual decomposition method from Hanley and Hoberg (2010) to separate the prospectus language into standard and informative content. The results indicate that the Regulation did not mechanically increase firm-specific disclosure once firm characteristics, year effects, and industry effects are accounted for. We also find no clear unconditional relationship between prospectus content and first-day underpricing in the Swedish setting. However, joint Wald F-tests of the content $\times$ regulation interaction terms indicate that the Regulation changed how the market prices disclosure – particularly for the Total Document ($F = 5.319$, $p < 0.001$) and the Risk Factors section ($F = 3.621$, $p < 0.01$) – suggesting that the Regulation affects pricing not only by changing what issuers disclose, but also by altering how investors interpret disclosure.

Keywords: IPO Underpricing, Prospectus Disclosure, EU Prospectus Regulation, Information Content, Swedish IPO Market

Authors:
Hedda Gummesson (25556)
Gustav Jäger (25743)

Tutor:
Paula Roth, Postdoctoral Fellow, Department of Finance

Examiner:
Ramin Baghai, Professor, Department of Finance

Bachelor Thesis
Bachelor Program in Business Economics
Stockholm School of Economics

© Hedda Gummesson och Gustav Jäger, 2026

1. Introduction

An IPO prospectus – the written disclosure document in which a firm describes its business, risks, and financial position before going public – is meant to reduce information asymmetry before the offer price is set. But the effect of regulation on this disclosure is not obvious. A new rule can make issuers provide more firm-specific information, or it can standardize the format in which existing information is presented. This paper studies this distinction using textual analysis of Swedish IPO prospectuses around the EU Prospectus Regulation ("Regulation"). We find that the Regulation does not robustly change the amount of informative content in IPO prospectuses, but it changes how the market prices standard and informative disclosure.

The textual analysis of our paper is based on Hanley and Hoberg (2010). They decompose prospectus text into standard, boilerplate content and informative, firm-specific content, finding that greater informative content produces more accurate offer prices and roughly 8% lower underpricing. We replicate and extend their methodology to the Swedish market to investigate how a different institutional setting, as well as a changing regulatory environment, affects price. The theoretical motivation for the link between prospectus content and price comes from Benveniste and Spindt (1989), who say that underpricing is the price an issuer pays to extract private information from investors during bookbuilding, so the more an issuer depends on bookbuilding for price discovery, the more money it must leave on the table. Prospectus quality directly mediates that dependence: when an issuer invests more time in pre-market due diligence and discloses more upfront, investors arrive at the bookbuilding already better informed, narrowing the information gap that underpricing must bridge.

Sweden presents an interesting setting for studying these questions. The country has for decades been a hub for entrepreneurial activity and company-building, and in terms of IPO activity, Sweden stands out as one of the most active markets in Europe – during 2025, Nasdaq Stockholm accounted for the highest share of global IPO transaction value among all European exchanges (Dalhäll, 2025). Sweden's combination of an active IPO market, a rather concentrated investment base, and smaller firm size may give rise to disclosure patterns that differ from those observed in larger markets such as the United States (US) (OECD, 2026). In July 2017, the EU Prospectus Regulation (EU) 2017/1129 ("Regulation") entered into force, stipulating in detail when and how a prospectus should be written, while also introducing the EU Growth Prospectus to reduce the regulatory burden for smaller issuers. For Sweden, as for all EU member states, this meant that IPOs completed after 21 July 2019, the full application date of the Regulation, had to comply with the new regulations. Whether prospectuses of firms operating under this regulatory framework look any different, and if this plays a role on the first-day prices, is the open empirical question throughout the paper.

The timing of this study is particularly relevant. As for global markets, IPO activity in Sweden declined sharply following the pandemic boom in 2021 and shrank even further during 2022 and 2023 (White & Case, 2024). The Swedish IPO market has begun to recover from these low levels, not fully reaching pre-pandemic levels, and pricing stands as a central consideration for issuers re-entering the market (Nasdaq, 2024). If a higher share of informative disclosure can lead to more accurate

pricing, the question of how to structure a prospectus can have an impact financially. This matters even more, given that producing prospectus content is costly – underwriter fees, which cover the drafting, researching, and verifying of prospectus content, represent the highest direct cost of going public, typically around 3–4% of proceeds in European markets (Abrahamson, Jenkinson, and Jones, 2011). Knowing which elements of prospectus content contribute to more accurate pricing is a way for issuers and their advisors to make better-informed decisions on where to allocate the costly consulting hours.

Along with the rise of IPOs, the rise of artificial intelligence (AI) is a motivation for this paper. AI is used increasingly in the financial industry, and language models' abilities to draft and compile reports are growing. As an example, the CEO of Goldman Sachs, David Solomon, has noted that AI can accurately compile around 95% of an IPO prospectus, reducing what previously required a six-person team two weeks of work to a matter of minutes (Financial Times, 2025). As AI tools take on a growing role in prospectus production, the question of which type of disclosure is associated with better pricing outcomes is relevant for how these tools should be designed and used.

Following Hanley and Hoberg (2010), we distinguish between the standard content of a prospectus, defined as language that mirrors recent and past industry IPOs, and the informative content, defined as disclosure unique to the issuing firm that is not explained by industry-wide patterns. Using this decomposition, we examine whether greater informative content leads to more accurate initial offer prices and lower underpricing, and conversely, whether greater standard content is associated with larger offer price revision and higher initial returns. We apply this framework to the Swedish market, using a sample set of 269 firms that are registered in Sweden and did their IPO on a Swedish stock exchange between January 2006 and February 2026.

We have structured the empirical analysis into two parts. First, we test if the EU Prospectus Regulation changed the informational composition of prospectuses (Prediction 1) by regressing standard and informative content scores – extracted for the whole document and six individual sections per prospectus – on dummies for the full Regulation and the EU Growth format. Second, we test if the different types of prospectus content predict first-day underpricing (Prediction 2), and if this result shifts under the new regulatory regime.

The results indicate that the Regulation did not mechanically increase the amount of firm-specific disclosure in Swedish IPO prospectuses. Once firm characteristics, year effects, and industry effects are accounted for, the Regulation itself is not associated with a significant change in either standard or informative content. Similarly, we do not find a clear unconditional relationship between prospectus content and first-day underpricing, meaning that the 8% reduction in underpricing documented by Hanley and Hoberg (2010) does not directly appear in the Swedish setting. However, the results suggest that the Regulation changed how the market prices disclosure. In the interaction regression, the joint Wald F-test rejects the null that the EU regime does not moderate the content-pricing relationship, particularly for the total document and risk factor sections. This implies that, while the Regulation did not substantially change what issuers disclose, it altered the pricing relevance of a given unit of disclosure. The strongest interpretation is therefore that the Regulation reshapes the information environment indirectly: not by increasing volume or

uniqueness of disclosure, but by changing how investors interpret standard and informative prospectus language. The EU Growth format shows a significant increase in informative content, though this is the only finding that does not hold up in our robustness checks and should be interpreted with caution. The remaining findings are stable across a range of robustness checks, including alternative error structures, PCA-based dimensionality reduction, and sub-period stability tests.

2. An Overview of the Context

This section provides the relevant context for the study, covering the related literature and the regulatory background that governs prospectus disclosure in the Swedish IPO market.

2.1 Related Literature

The introduction of the Regulation in July 2019 enables a quasi-experimental design, because the regulatory implementation itself can be treated as an exogenous shift in the information environment rather than a choice made by individual issuers. A well-identified example of how the literature exploits regulatory implementation in this way is Chang, Ljungqvist, and Tseng (2023), who use the staggered rollout of EDGAR to show its consequences on analyst behavior, finding that mandatory disclosure constrains their strategic forecasting, making predictions less optimistic and more accurate. We adopt a parallel approach, using the EU/ESMA regulatory shift as an exogenous event to look at how it altered the informational composition of prospectuses and the pricing of disclosure.

When assessing IPO success, much of the literature focuses on first-day prices. Carter and Manaster (1990) look at the underwriter reputation implication on first-day prices; Megginson and Weiss (1991) look at the implication of venture backing on first-day prices; Aggarwal, Prabhala, and Puri (2002) look at whether institutional demand can serve as a proxy for first-day prices. Our study similarly uses first-day price, comparing it to the price stated in the IPO prospectus, to measure pricing accuracy – defined as the initial offer price relative to the price set by the market on the first day of trading.

To analyze the textual content of the IPO prospectuses, we follow Hanley and Hoberg (2010), who ask whether issuers face a tradeoff between investing in pre-market due diligence and relying on bookbuilding to price their issue.[1] They decompose prospectus text into two components: standard content, defined as language that mirrors prior IPOs within the same industry, and informative content, defined as disclosure unique to the issuing firm that is not explained by industry-wide patterns, using the latter as a proxy for pre-market effort expended. Their central finding is that greater informative content leads to more accurate initial offer prices, smaller price revisions between the initial filing and the final offer, and approximately 8% lower underpricing. The opposite holds for standard content – a higher share of boilerplate language is associated with larger upward price revisions and greater first-day returns.

[1] Pre-market due diligence involves the production of information about the firm ahead of the offering, which feeds into the prospectus content. Bookbuilding is the process by which bids from investors are collected, typically by the underwriter, to determine the offer price.

This tradeoff is grounded in Benveniste and Spindt (1989), who show that underpricing is the payment an issuer must make to extract private information investors have about the value of the issuer during bookbuilding. The more an issuer depends on bookbuilding for price discovery, the more it must leave on the table to compensate investors for sharing their private information. Prospectus quality directly mediates this dependence: when an issuer discloses more upfront, investors arrive at bookbuilding already better informed, narrowing the information gap that underpricing must bridge. Cornelli and Goldreich (2001) provide evidence of this mechanism, showing that bookbuilding allocations are used strategically to reward investors who submit more informative bids and investors with repeated participation in the IPO market. It is in this setting that an informative prospectus can reduce the dependence on costly investor information during bookbuilding, as Hanley and Hoberg (2010) show by finding that standard content leads to higher underpricing. In other words, relying on bookbuilding instead of pre-market due diligence leads to higher underpricing.

However, underpricing can also arise for reasons other than incentivizing information revelation during bookbuilding. An alternative explanation comes from Rock (1986), who shows that when some investors are better informed than others, uninformed investors face a winner's curse: they are more likely to receive shares in unattractive offerings and to be crowded out of attractive ones. Therefore, to keep uninformed investors willing to participate, issuers must offer shares at a discount. The relation between prospectus content and underpricing may depend on the IPO pricing mechanism. If underpricing mainly compensates informed investors for revealing information during bookbuilding, the informative content of a prospectus should matter directly for pricing accuracy. If underpricing instead reflects a discount to increase participation, the link between informative prospectus content and first-day return may be weaker.

The studies discussed above establish the theoretical and methodological foundation of our analysis, but they are rooted primarily in the US market context. Replicating Hanley and Hoberg (2010) in a Swedish setting can imply a different empirical result, as market structure, investor base, and regulatory environment differ. Sweden is a small open economy with a highly developed equity market relative to its size, with new listings taking place on the regulated markets Nasdaq Stockholm and Nordic Growth Market (NGM), as well as on multilateral trading (MTF), including First North, Spotlight Stock Market, and NGM SME (SEB, 2026). Participation in Swedish IPOs is dominated by a concentrated domestic investor base – traditional institutional investors, including investment funds, pension funds, and insurance companies, collectively owned more than half of all listed Swedish companies at the end of 2024, with a small number of large repeat players, such as the national pension funds and major family-owned investment firms, operating continuously across the same industries and intermediaries (OECD, 2026). These features make Sweden an interesting setting to test whether the content-pricing relationship documented in the US setting holds in a setting where price discovery may rely less on prospectus disclosure alone and more on the interactions and relations of issuers, underwriters, and investors.

Against this backdrop, this paper makes several contributions to the literature. First, it offers a replication of Hanley and Hoberg's (2010) textual methodology outside the US market, testing whether the distinction between standard and informa-

tive prospectus content has the same pricing implications in a different institutional setting. Second, we go a step further than the original framework by including the regulatory environment and treating it as something to be studied rather than a background condition – using the introduction of the EU Prospectus Regulation in 2019 as a natural experiment to examine whether mandatory disclosure requirements change what issuers write and how the market prices it. To the extent that we find a relationship between informative content and first-day prices, the logic of Benveniste and Spindt (1989) and Hanley and Hoberg (2010) provides a way to interpret it: issuers who disclose more upfront reduce the information gap between themselves and investors, which in turn reduces the need to underprice. In this way, the paper extends an established methodology to a new market while adding a regulatory dimension that the original study did not address.

2.2 Regulatory Background

The regulatory framework governing prospectus disclosure in Sweden has undergone substantial transformation over the period covered by our sample. Prior to 2005, Swedish prospectus rules consisted of multiple different laws,[2] guidelines from the Swedish Financial Supervisory Authority ("Finansinspektionen") and industry self-regulation, such as exchange rulebooks and market practice (Andersson, 2009). These parallel regimes applied different concepts and thresholds for when a prospectus was required and what it was obligated to disclose, resulting in prospectuses that varied substantially in both structure and content.

For Sweden, this domestic heterogeneity was resolved with the implementation of the EU Prospectus Directive (2003/71/EC) ("Directive") – adopted at the EU level to harmonize prospectus requirements across member states – through SFS 2005:833[3], which came into effect on 1 January 2006 (Swedish Parliament, 2005). The reform consolidated prospectus rules into a single, comprehensive chapter of the Act on Trading in Financial Instruments (LHF) 1991:980 (Swedish Parliament, 1991) and assigned Finansinspektionen (FI) the responsibility for prospectus approval. It also introduced a uniform materiality standard, requiring disclosure of all information necessary for investors to assess the issuer's financial position and prospects. The detailed content requirements were further specified by the European Commission (2004), which prescribed annexes covering historical financial information, risk factors, business description, governance, ownership structure, and dividend policy, among other items. This created the baseline structure for all Swedish IPO prospectuses on regulated markets between 2006 and 2019, and it is the regime under which the first part of our sample was produced.

A following, more fundamental shift occurred with the adoption of the EU Prospectus Regulation (European Parliament and Council, 2017), which replaced the Directive and took full effect on 21 July 2019. The motivation behind the Regulation was based on evaluations of the Directive's finding that prospectuses had become excessively long and increasingly built on generic, compliance-oriented language, which was more to manage issuers' liabilities than to provide meaningful

[2]The laws included Aktiebolagslagen (1975:1385), Act on Stock Exchange and Clearing Operations, and the Act on Trading in Financial Instruments (LHF) 1991:980.

[3]SFS (Svensk författningssamling) is the official Swedish statute book in which all Swedish laws and ordinances are published and numbered chronologically by year and sequence number. A reference such as SFS 2019:414 thus denotes the 414th entry in the statute book for 2019.

information to investors. To achieve this, the Regulation introduced clearer and more standardized templates for the ones writing prospectuses, replacing the more flexible structure of the Directive with a mandatory format. Risk factor presentation was the section most affected – under the new, issuers were required to limit risk factors only to those that are specific and material to the issuer and its securities, to organize them into a limited number of categories, and to rank them within each category by the estimated likelihood and potential impact. The summary section was also subject to new requirements, now having to be divided into four sections covering the issuer, the securities, the offer, and the key risks, at a maximum length of seven pages. The Regulation also introduced proportionate prospectus formats – the EU Growth Prospectus for SME issuers and a simplified format for secondary issuances – intended to reduce the disclosure burden for smaller companies without sacrificing investor protection.

With the Regulation in place, Sweden replaced the prospectus chapters of LHF with SFS 2019:414 (Swedish Parliament, 2019) so that after 2019 the content requirements are found in the Regulation and its delegated acts. SFS 2019:414, among other things, specifies 2.5 million euros as Sweden's offering threshold below which an EU-compliant prospectus is not required (the Regulation permits EU member states to set this anywhere between 1 and 8 million euros). The purpose of this threshold relates to the venues on which securities are typically listed in Sweden. For admission to a regulated market within the meaning of MiFID II[4] – such as Nasdaq Stockholm – the Regulation mandates a full EU prospectus unconditionally under Article 3(3), regardless of the size of the offering, meaning the 2.5 million euros threshold is simply irrelevant here. For listings on a MTF, admission does not by itself trigger the Regulation's prospectus obligation; what matters instead is whether the listing is accompanied by a public offer of securities. If that public offer raises less than 2.5 million euros, no prospectus is required at all; if it raises more, a prospectus is required. However, issuers qualifying as SMEs under the MiFID II definition[5] do not need to write a full prospectus, but can under Article 15 satisfy this obligation by writing a lighter EU Growth Prospectus. Beyond the threshold rules, SFS 2019:414 also defines the liability framework for prospectus content and keeps FI as the supervisory authority. FI may require issuers to supplement disclosures where deemed necessary for investor protection, making it a natural data source for us, as every prospectus in the sample has the same level of regulation, quality, and completeness control.

[4]MiFID II (Markets in Financial Instruments Directive II) is the EU legislative framework governing financial markets across member states, in force since January 2018. Among other things, it defines what constitutes a "regulated market" – a formally authorized trading venue subject to strict requirements on transparency, oversight, and investor protection.

[5]Article 2(f) of European Parliament and Council (2017) states that the issuer qualifies as an SME if, according to its last annual or consolidated accounts, it meets at least two of the following three criteria: (1) an average number of employees during the financial year of fewer than 250; (2) a total balance sheet not exceeding €43 million; and (3) an annual net turnover not exceeding €50 million. Alternatively, an issuer qualifies if its average market capitalization was less than €200 million, based on end-year quotes for the previous three calendar years.

3. Data Description

To construct our sample of IPO prospectuses from Swedish firms, we employed a web crawling algorithm on FI's website. We extracted three separate datasets from FI: (1) the prospectus archive, with 1,510 prospectuses; (2) the current prospectus registers for EU Growth prospectuses, with 500 prospectuses; and (3) the current prospectus registers for full prospectuses with, 219 prospectuses. The prospectuses available from FI range from October 2006 to January 2026, which sets the outer bounds for the sample period. The files were saved using a standardized naming convention of firm name followed by prospectus date (`FirmName_Date.pdf`). We have included IPOs from regulated and unregulated markets, as both produce a full EU prospectus: regulated markets in every case, and unregulated markets when the offer exceeds 2.5 million euros, and the issuer qualifies as an SME under the MiFID II definition, as explained in Section 2.2.

To identify true IPO transactions, we collected a dataset from S&P Capital IQ, containing all completed IPO transactions in Sweden between 2006 and 2026, resulting in 725 observations. This dataset includes parameters such as firm name, transaction date, and the percentage change in stock price one day after the transaction, which are used in the main analysis. The transaction date is particularly important since, after regulatory approval, firms have 365 days to issue shares but are not obligated to do so – and our analysis requires confirmed transactions only. All algorithmic matches between the name on FI's website and the transaction name from S&P Capital IQ were subsequently verified manually to ensure accuracy. For transactions lacking a corresponding prospectus in the FI archive, we attempted to source the correct document manually. In the majority of these cases, we did not find the full prospectus but an information memorandum for the firm, which is a common lighter document to present when exempted from issuing a full prospectus. The information memoranda are not included in the sample, as they are not official prospectuses and do not contain the structured disclosure content required for our analysis. This matching and filtering process left us with 409 transactions with a corresponding prospectus. The significant attrition from the sample of around 700 prospectuses is primarily explained by the number of simplified prospectuses ("Förenklat prospekt") in the raw data. Like the memoranda, these prospectuses are a lighter version of a full prospectus for firms conducting secondary offerings, employee share schemes, or certain merger-related share issuances. Because these documents do not follow the same disclosure structure as full IPO prospectuses and often do not represent first-time public offerings, they are excluded from our analysis.

We then compare the prospectus dates, taken from the file name, with the transaction date from the S&P Capital IQ dataset, to ensure that each prospectus is available within a span of 365 days from the offering, the legal time limit within which the firm must issue after approval of their prospectus. This reduces the sample to 380 prospectuses, which are subsequently converted from PDF to HTML using Claude Opus 4.6 (Anthropic) to enable text comparison. Due to the structural differences across documents – including variation in formatting, embedded tables, and scanning quality – 11 prospectuses could not be converted reliably, leaving a sample of 269 prospectuses.

For the 269 final IPO prospectuses, we compiled a dataset, extracted from S&P

Capital IQ, that includes transaction date, one-day price change, industry group, topics, agents (main underwriter, legal counsel and auditor), private equity affiliation, price of the IPO, exchange, gross proceeds and shares offered. The dataset also includes the prospectus date, taken from the file name of each document, and two dummy variables for EU Growth and EU Regulation prospectuses. The EU Growth dummy equals 1 where FI's document type field ("Dokumenttyp") classifies the prospectus as a "Tillväxtprospekt", while the EU Regulation dummy equals 1 for all prospectuses issued after 21 July 2019. For the missing data points in the dataset, as well as to gather the expenses for the issue, the prospectus share price, and filing size, we manually reviewed the prospectuses. The share price and the price of the IPO helped us compute the bookbuilding price change (ΔP). Following this, the data regarding agents is manually summarized by only keeping the main company and not any department-specific information (e.g. DNB Markets goes into DNB Carnegie). Market data regarding OMX30 are taken from Capital IQ as the daily closing price and then denominated as the 30-day average leading up to the IPO date. The tech dummy is defined through the company's topics and industry. Any missing value after this is manually checked by looking through relevant websites, such as Bolagsverket, for company incorporation date. After these steps, the final sample consists of 269 IPOs spanning the period 2007–2025[6].

4. Text Measurement and Content Decomposition

This section explains how the textual content of each IPO prospectus is transformed into measurable variables. We start by defining the key pricing, firm-level, and text-based variables used in the analysis. We then describe how each prospectus is divided into six core sections and converted into normalized word vectors. Finally, we explain how these word vectors are decomposed into standard and informative content, which form the central disclosure measures used in the empirical analysis in Section 5.

4.1 Variable Description

With the data in our master dataset, we compute several different variables, all necessary for the research question and control variables we aim to examine. Panel A consists of three main variables: price adjustment (ΔP) declares the direction of the adjustment between the price found in the prospectus and the price on the first day of trading; absolute price adjustment ($|\Delta P|$) is the magnitude of that price direction; and the initial return (IR) is the underpricing, meaning the percentage change from the offer price to the first-day closing price in the secondary market.

In Panel B we find variables important for the pricing data, as well as characteristics of the firm, its underwriter, and the market the firm operates in. These variables and their individual descriptions are found in Appendix A.

As for the prospectus text variables, Panel C, the total length of the document is noted as total words in document. The documents are then split into sections to see each part of the document's importance to the price adjustment. These six sections are listed and explained in Section 4.3. For comparison purposes, these are listed as

[6]No viable IPO prospectuses were identified for 2006 and 2026 following our sample construction, leaving an effective sample period of 2007-2025.

individual shares of the total document. Panel D, word vectors, are the root words per section, being words that appear five or more times in all prospectuses, after removing articles, conjunctions, and personal pronouns.

Table 1: Summary Statistics

	Mean	Std. dev.	Minimum	Median	Maximum
Panel A: Price variables					
Price adjustment (ΔP)	0.085	0.624	−0.875	0.000	8.636
Absolute price adjustment ($\lvert\Delta P\rvert$)	0.110	0.620	0.000	0.000	8.636
Initial return (IR)	0.121	0.302	−0.872	0.066	1.756
Panel B: IPO characteristics					
IPO size at filing (MSEK)	1,015.9	2,998.3	5.555	195.8	38,525.9
Gross proceeds (MSEK)	933.6	2,306.9	5.250	195.2	23,000.0
Firm age (years)	10.425	12.634	0.066	7.439	115.9
Tech dummy	0.361	0.481	0.000	0.000	1.000
PE dummy	0.145	0.353	0.000	0.000	1.000
Financial sponsor dummy	0.227	0.420	0.000	0.000	1.000
Regulated market dummy	0.792	0.407	0.000	1.000	1.000
EU regulation dummy	0.528	0.500	0.000	1.000	1.000
EU growth dummy	0.201	0.401	0.000	0.000	1.000
Lead UW market share	0.161	0.223	0.000	0.041	1.000
Pre-file OMX30 return	0.005	0.043	−0.272	0.007	0.101
Panel C: Prospectus text					
Total words in document	47,324.6	28,553.0	8390.0	38,916.0	180,945.0
Summary / total document	0.048	0.102	0.000	0.019	0.772
Risk factors / total document	0.126	0.082	0.000	0.118	0.707
Background & motive / total document	0.065	0.109	0.000	0.026	0.756
Business description / total document	0.260	0.206	0.000	0.223	0.857
Financial information / total document	0.340	0.209	0.000	0.350	0.844
Other content / total document	0.162	0.110	0.000	0.145	0.723
All six sections	1.000	0.000	0.999	1.000	1.000
Panel D: Word vectors (root words only)					
Total document	12,274.2	7,564.9	297.0	9,663.0	46,145.0
Summary	608.9	1,187.0	7.0	264.0	11,354.0
Risk factors	1,761.7	1,257.8	5.0	1,464.5	8,802.0
Background & motive	576.3	1,047.2	4.0	250.0	6,866.0
Business description	2,692.1	2,122.5	6.0	2,481.0	19,951.0
Financial information	5,153.5	4,635.0	93.0	3,935.0	33,622.0
Other content	1,867.0	1,261.8	15.0	1,677.0	8,317.0

Notes: Summary statistics for 269 Swedish IPO prospectuses (2007–2025). Price adjustment (ΔP) is the return from the prospectus price to the IPO offer price (of which 77% are zero). Initial return (IR) is the percentage change from the offer price to the first-day closing price. Firm age is IPO year minus founding year. Lead UW market share is the lead underwriter's gross-proceeds share in the calendar year of the IPO. Word vector statistics are root words after Swedish stemming.

An interesting finding from Panel A in Table 1 is the price adjustment (ΔP): 77% of observations are exactly zero, yielding a mean of 0.085 but a median of zero. This reveals a feature of the Swedish IPO market that is different from the US setting studied by Hanley and Hoberg (2010), where bookbuilding-driven price revisions are both common and play a central role in the price discovery process. Sweden seems to rely more on fixed pricing, and given this, we will place greater emphasis on the initial return as the primary measure of underpricing in our analysis.

The initial return has a mean of 12.1% and a median of 6.6%, which, combined with the 77% share of zero-price adjustments, suggest that IPOs priced within a range systematically set their offer price too low – the market consistently prices them higher on the first day of trading. This may be a consequence of Sweden's high share of fixed price offerings, not enabling the underwriters to revise prices along with demand. It can also be a consequence of Sweden's concentrated investor base, implying that the larger informed investors cluster around a few offers, and underpricing is there to incentivize the uninformed to participate in the less prominent offerings. Further, the right-skewed distribution, with a maximum initial return of 175.6%, suggests the presence of a few highly speculative listings.

Turning to Panel B, the IPO size at filing has a mean of 1,016 MSEK but a median of only 196 MSEK, indicating a heavily right-skewed distribution driven by a small number of large transactions. Gross proceeds follow a similar pattern, with a mean of 934 MSEK and a median of 195 MSEK. Firm age averages 10.4 years at the time of listing, with a median of 7.4 years – compared to the US market of Hanley and Hoberg (2010) with a mean age of 13.4 years. This suggests that Swedish IPO firms are relatively young compared to a typical US sample, a pattern likely driven by Sweden's early-stage technology firms.

The EU Regulation dummy indicates that 52.8% of prospectuses fall under full EU regulatory requirements implemented in July 2019, and the EU Growth dummy captures that 20.1% of all observations adopted the lighter-touch EU Growth Prospectus format introduced the same year. This implies that approximately 38% of post-Regulation IPOs opted for the simplified prospectus regime – a distributional split with direct implications for our interaction results, as firms face meaningfully different disclosure requirements depending on the format chosen. Lead underwriter market share has a mean of 16.1% but a median of just 4.1%, suggesting a concentrated market with a few dominant major banks and many smaller ones. This skewness implies that the underwriter market share variable captures large differences in quality and reputation across firms.

Panel C describes the textual content of the prospectuses. The average document contains 47,324 words (median 38,916), though variation is substantial, ranging from as few as 8,390 to as many as 180,945 words, suggesting considerable heterogeneity in issuers' disclosures. This breadth of variation is consistent with what Hanley and Hoberg (2010) observe in US prospectuses, where document length and informational composition differ markedly across firms and market conditions. Within the document, Financial Information constitutes the largest proportional share at 34.0% on average, followed by Business Description at 26.0% and Risk Factors at 12.6%. The Summary section, by contrast, accounts for only 4.8% of total document length. All six sections add up to one, showing that the splitting of the text is consistent across documents.

Panel D reports word vector statistics constructed from root words only, following Swedish Snowball stemming. The total document has a mean of 12,274 stemmed words (median 9,663), substantially fewer than the raw word count in Panel C. The Financial Information and Business Description sections again dominate in absolute size, with means of 5,153 and 2,692 root words respectively. The Risk Factors section sits in the middle, with an average of 1,762 root words. In contrast, the Background & Motive section – which is important for understanding why firms go public – is much shorter, averaging 576 root words. However, it varies considerably across firms

(Std. Dev. = 1,047), suggesting that companies differ widely in how they explain their reasons for going public.

4.2 Research Design

We aim to explore how the information content has changed after the EU Regulation went into effect in 2019, and subsequently, if this has influenced the pricing of IPOs on the Swedish market. To do this, we recreate the text analysis model employed by Hanley and Hoberg (2010) to divide each IPO prospectus into standard and informative content.

By looking at the information content before and after the Regulation came into effect, we can examine whether the Regulation had an effect on the Swedish IPO market. From this, we expect that:

Prediction 1: The EU Prospectus Regulation will be associated with an increase in informative content, particularly in the Summary and Risk Factors sections.

Prediction 2: A higher degree of informative content will be associated with lower initial returns, as greater pre-market disclosure reduces information asymmetry between issuers and investors.

4.3 Text Measures

Each IPO prospectus was divided into six core sections. These six sections were chosen to reflect the major building blocks of a Swedish IPO prospectus as mandated under European Commission (2004) and the Regulation. Each section serves a distinct informational purpose and targets different aspects of the issuer and the offering, meaning that content similarity and informational properties are likely to vary systematically across them. Analyzing sections separately, therefore, allows us to assess whether the effects of the Regulation were concentrated in specific parts of the document rather than distributed uniformly. A detailed description of the content and role of each section is provided in Appendix B.

For each IPO i, we stored the HTML prospectus text into seven different word vectors, denoted:

- $\widetilde{\mathbf{w}}_{\text{tot},i}$ containing the *entire document*
- $\widetilde{\mathbf{w}}_{\text{sum},i}$ containing the *Summary* section
- $\widetilde{\mathbf{w}}_{\text{risk},i}$ containing the *Risk Factors* section
- $\widetilde{\mathbf{w}}_{\text{B\&M},i}$ containing the *Background & Motive* section
- $\widetilde{\mathbf{w}}_{\text{bus},i}$ containing the *Business Description* section
- $\widetilde{\mathbf{w}}_{\text{fin},i}$ containing the *Financial Information* section
- $\widetilde{\mathbf{w}}_{\text{other},i}$ containing the *Other* section

All these vectors are based on the same world root vector, containing 3,496 unique word roots. We use word roots rather than actual words to capture variations in words that, in essence, are the same. Furthermore, certain types of words have been excluded, such as articles, conjunctions, and personal pronouns (for more information about the creation of the word root vector, see Appendix D). By construction, all seven vectors for a given IPO i have the same length, 3,496, where each element of the vector shows how many times that specific root word exists in the text/section. All vectors are then normalized in order to only capture the relative occurrence of a word, disregarding how long the text/section as a whole is. This normalization is important for comparison, as EU Growth prospectuses follow

a similar disclosure structure to full prospectuses, but in a shorter format.

4.4 Common IPO Content

With the normalized word vectors constructed, we can look at what drives similarities and differences in prospectus content across our sample. To do this, we compute the cosine similarity between every possible pair (i, j) of IPOs, both for the total document and for each of the six sections separately. The cosine similarity between two word vectors is bounded between 0 and 1, where a value of 1 indicates that the two documents have an identical relative word distribution, and a value of 0 indicates that they share no words at all. With a sample of 269 IPOs, this yields 19,503 unique pairs per word vector. We then regress these pairwise similarities on a set of variables describing how similar two IPOs are in terms of participant variables, market and timing, and firm characteristics, with standard errors clustered by IPO i to account for each IPO appearing in multiple pairs:

$$\text{CosineSim}_{ij,s} = \alpha + \boldsymbol{\beta}_1 \mathbf{P}_{ij} + \boldsymbol{\beta}_2 \mathbf{M}_{ij} + \boldsymbol{\beta}_3 \mathbf{D}_{ij} + \boldsymbol{\beta}_4 \mathbf{C}_{ij} + \varepsilon_{ij} \tag{1}$$

where $\text{CosineSim}_{ij,s}$ is the cosine similarity between IPO i and IPO j for section s. Participant variables $\mathbf{P}_{ij}$ include indicators for Same Lead Underwriter, Same Underwriter Counsel, and Same Auditor. Market and timing variables $\mathbf{M}_{ij}$ include Same Industry Group and Same Transaction Window, where the window is set to 90 days following Hanley and Hoberg (2010). Dummy variables $\mathbf{D}_{ij}$ include indicators for Both Tech, Both PE-backed, Both EU Regulation, Both EU Growth, and Both Regulated Market. Continuous variables $\mathbf{C}_{ij}$ include the absolute difference in firm age (in years), the absolute difference in prospectus year, and the absolute difference in log filing size. Standard errors are clustered by IPO i, to account for each IPO appearing in $N - 1 = 268$ pairs.

A positive coefficient on a given variable therefore means that sharing that characteristic is associated with more similar prospectus content, while a negative coefficient indicates the opposite. The regression thus tells us what factors systematically explain why two prospectuses look alike – in other words, it reveals the sources of standardized content in our sample. This is an important step, as it shows that there are clear and systematic factors behind content similarity, which supports and motivates the next step, where each prospectus is decomposed into standard and informative components.

Looking at the participant variables in Table 2, the Same Lead Underwriter coefficient is negative for the whole document (-0.016^{*}), suggesting that sharing an underwriter does not, on average, produce more similar prospectuses overall. However, the same underwriter is associated with significantly more similar Financial Information (0.057^{***}) and Other content (0.021^{***}), while being more different in sections such as Risk Factors (-0.054^{***}) and Business Description (-0.024^{**}). This pattern is theoretically interesting – underwriters appear to apply common financial templates and accounting conventions across clients, producing convergence in disclosure-heavy sections, while simultaneously tailoring firm-specific narrative sections to each issuer.

Turning to market and timing variables, Same Industry Group is positive and significant across most sections, consistent with Hanley and Hoberg's (2010) finding that industry-level convergence is a primary driver of prospectus similarity, and

validating our use of industry content in the standard content decomposition in Section 4.5.

Among the dummy variables, the EU Regulation coefficients produce the most directly relevant findings in terms of the regulatory change in Table 2. Both EU Regulation is strongly positive for the Summary section (0.114^{***}), indicating that post-Regulation summaries became more standardized – consistent with the Regulation's goal of a short, structured summary format. At the same time, Both EU Regulation is significantly negative for Risk Factors (-0.051^{***}), suggesting that risk factor disclosure became more firm-specific following the Regulation. This result is consistent with the Regulation's intent to move issuers away from generic risk lists toward identifying only risks that are specific and important to the firm and its offering. For EU Growth, the results are also in line with the Regulation's intent: Both EU Growth is negative for Financial Information (-0.101^{***}) and Other content (-0.106^{***}), indicating that EU Growth prospectuses produce less similar financial sections across pairs. This difference likely reflects the lighter disclosure requirements of the EU Growth Prospectus, giving issuers more flexibility in what information to include in the large Financial Information section, as seen in Table 1.

The R^2 values range from 5.4% to 20.7%, which are modest but not surprising given the nature of the analysis. The goal of the regression is not to fully explain why prospectuses are similar, but to show that some clear and consistent factors influence this similarity. The remaining variation – the content that cannot be explained by participant characteristics, industry, timing, or regulatory format – is precisely what the standard and informative content decomposition in the next step is designed to capture.

Table 2: Sources of Prospectus Content Similarity

	Total	*Sum.*	*Risk*	*B & M*	*Bus.*	*Fin.*	*Other*
Participant Variables							
Same Lead Underwriter	−0.0159***	0.0106	−0.0537***	−0.0028	−0.0241**	0.0566***	0.0211***
	(−3.011)	(0.823)	(−6.935)	(−0.369)	(−2.613)	(7.974)	(2.867)
Same Underwriter Counsel	0.0032	−0.0094	0.0280***	0.0078	−0.0147*	0.0197***	0.0085*
	(0.795)	(−0.892)	(3.671)	(1.127)	(−1.881)	(4.354)	(1.792)
Same Auditor	0.0019	0.0137	−0.0010	0.0131*	−0.0140**	0.0059	0.0103*
	(0.602)	(1.636)	(−0.202)	(1.862)	(−2.398)	(1.534)	(1.766)
Market & Timing							
Same Industry Group	0.0187***	0.0172*	0.0323***	0.0197**	0.0540***	0.0053	−0.0024
	(4.318)	(1.710)	(4.599)	(2.211)	(5.672)	(1.291)	(−0.402)
Same Transaction Window (3m)	−0.0052	−0.0161*	−0.0033	−0.0034	0.0254***	−0.0103**	−0.0163***
	(−1.197)	(−1.885)	(−0.454)	(−0.456)	(3.458)	(−2.339)	(−2.708)
Dummy Variables							
Both Tech	0.0094*	0.0130	0.0148*	−0.0222**	0.0331***	−0.0018	0.0070
	(1.676)	(0.909)	(1.655)	(−2.461)	(3.024)	(−0.247)	(0.895)
Both PE-backed	0.0059	0.0683**	−0.0277	−0.0193	0.0453***	0.0473***	0.0081
	(0.420)	(2.249)	(−1.389)	(−1.106)	(2.631)	(4.207)	(0.610)
Both EU Regulation	0.0024	0.1140***	−0.0505***	0.0170	0.0307**	0.0112	0.0052
	(0.332)	(6.052)	(−4.528)	(1.418)	(2.260)	(1.330)	(0.664)
Both EU Growth	0.0181	0.0156	0.0297	0.0521***	0.0897	−0.1014***	−0.1056***
	(1.514)	(0.705)	(1.013)	(3.286)	(1.621)	(−5.021)	(−8.382)
Both Regulated Market	−0.0029	0.0014	−0.0133	−0.0087	−0.0096	0.0195*	0.0100
	(−0.562)	(0.117)	(−1.376)	(−1.085)	(−0.565)	(1.892)	(1.404)
Continuous Variables							
|Firm Age Difference| (years)	−0.0003***	−0.0009***	−0.0003	−0.0006***	0.0003	0.0001	0.0003***
	(−2.956)	(−2.650)	(−0.966)	(−2.934)	(1.328)	(0.976)	(2.832)
|Prospect Year Difference|	−0.0031***	−0.0053***	−0.0085***	−0.0079***	0.0006	−0.0031***	−0.0091***
	(−2.644)	(−2.655)	(−6.209)	(−3.037)	(0.309)	(−2.829)	(−3.576)
|Log Filing Size Difference|	−0.0278***	0.0013	−0.0326***	−0.0165***	−0.0153***	−0.0271***	−0.0075***
	(−12.915)	(0.457)	(−10.385)	(−6.439)	(−5.667)	(−16.116)	(−4.316)
Observations	19,503	16,653	19,306	18,145	16,653	19,503	19,503
R^2	0.168	0.068	0.110	0.069	0.054	0.207	0.088

Notes: OLS regressions (Equation 1) with standard errors clustered by IPO i. The dependent variable is the cosine similarity between prospectus pair $(i, j) \in [0, 1]$. All specifications include year and industry fixed effects. Remaining columns use section-level vectors. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

4.5 Informative versus Standard Content

In order to assess whether prospectuses contained more information following the Regulation, we decompose each prospectus word vector, $\widetilde{\mathbf{w}}_{\text{tot},i}$, into its standard and informative components. Following Hanley and Hoberg (2010), we define "standard" content as content driven by information found in prior IPOs, and "informative" content as content that is unique to a given IPO.

Standard content is estimated using two sources. The first is *recent content* (rec): the average normalized word vector of all IPOs filed within 90 days prior to IPO i's filing date. This captures boilerplate language related to current market conditions, deal structures, and regulatory disclosures that tend to cluster in time. The second source is *industry content* (ind): the average normalized word vector of same-industry IPOs filed more than 90 days before IPO i, with no upper bound. The 90-day cutoff ensures that the two sources do not overlap. As shown in section 4.4, both components have a strong systematic relation to the wording of prospectuses, consistent with the findings of Hanley and Hoberg (2010).

Additionally, for each IPO i with K recent IPOs and P same-industry past IPOs, we define:

$$\widetilde{\mathbf{w}}_{\text{rec},i} = \frac{1}{K} \sum_{k=1}^{K} \widetilde{\mathbf{w}}_{\text{tot},k} \tag{2}$$

$$\widetilde{\mathbf{w}}_{\text{ind},i} = \frac{1}{P} \sum_{p=1}^{P} \widetilde{\mathbf{w}}_{\text{tot},p} \tag{3}$$

We then estimate the following first-stage regression (without an intercept) for each IPO i, where each observation corresponds to one word root, for a total of 3,496[7] observations:

$$\widetilde{\mathbf{w}}_{\text{tot},i} = a_{\text{rec},i}\, \widetilde{\mathbf{w}}_{\text{rec},i} + a_{\text{ind},i}\, \widetilde{\mathbf{w}}_{\text{ind},i} + \boldsymbol{\varepsilon}_i \tag{4}$$

Standard content for IPO i is then defined as the sum of the two fitted coefficients:

$$a_{\text{standard},i} = \hat{a}_{\text{rec},i} + \hat{a}_{\text{ind},i} \tag{5}$$

This scalar measures the loading of IPO i's prospectus on the word distribution of prior IPOs, that is, the proportion of its content that is attributable to standard, boilerplate language. A higher value indicates that more of the prospectus mirrors what has appeared in recent and industry-comparable filings.

Informative content is defined as the sum of the absolute values of the residuals from eq. (4):

$$a_{\text{informative},i} = \sum_{w} \left| \varepsilon_{i,w} \right| \tag{6}$$

This captures word-by-word deviations from what would be predicted by the standard content model alone, i.e. the firm-specific language that the issuer adds beyond any boilerplate. The use of absolute values ensures that both over- and under-representation of words relative to the standard contribute positively to the

[7] 19 word roots were not found in any of the 269 prospectuses and were therefore dropped from the vocabulary, reducing the word vector dimension from 3,515 to 3,496.

informative content measure, so that an unusually low mention of a root word is treated as equally informative as an unusually high one.

Standard and informative content are estimated separately for each prospectus section (Summary, Risk Factors, Background & Motive, Business Description, Financial Information, and Other) using only the word vectors from that section, so that section-level coefficients are based on comparisons to the corresponding sections of recent and industry IPOs. This allows us to assess whether the EU Prospectus Regulation's effects were concentrated in specific parts of the document, such as the Summary or Risk Factors sections, which were subject to significant regulatory changes. The resulting descriptive statistics for both standard and informative content, across all sections, are reported in Table 3.

In Panel A, informative content has a higher mean (0.585 for the total document) but also a higher standard deviation – the residual component varies more across firms than the peer-loading component. This means that the informative content measure might be picking up firm-specific heterogeneity.

In Panel C, standard content increases significantly post-Regulation for B&M ($+0.112^{***}$), Bus ($+0.138^{***}$), Fin ($+0.053^{**}$), and Other ($+0.042^{**}$). This means prospectuses became more standardized in almost every section after 2019, which is partly in line with the Regulation's intent of being standardized in structure, but firm-specific in content. Informative content shows a significant negative shift only for the total document (-0.028^{**}) and Other (-0.064^{**}), suggesting uniqueness declined at the margin. This is an important finding since the Regulation intended to make prospectuses more informative, but the descriptive evidence suggests the opposite occurred, at least in terms of textual uniqueness relative to peers.

The Risk Factors section shows no significant change in either standard or informative content post-Regulation, which is surprising given the Regulation's specific focus on improving risk factor disclosure. One interpretation is that the qualitative improvement – better categorization and prioritization – does not register in a word-vector similarity measure, which captures vocabulary overlap rather than logical structure.

The Summary and Risk Factors sections show no significant change despite the Regulation mandating a structured short summary and shorter, more specific risk factors. This might reflect that summaries were already relatively standardized, or that the new format requirements changed structure without changing vocabulary. Our word-vector approach also disregards words that are not in the world root list, which means that it cannot capture changes in text where there are few of these words.

Table 3: Standard vs. Informative Content

Content	Stat	*Total*	*Sum.*	*Risk*	*B & M*	*Bus.*	*Fin.*	*Other*
Panel A: Full sample								
Standard	Mean	1.001	0.882	0.979	0.919	0.906	0.972	0.973
	Std. dev.	0.152	0.496	0.237	0.277	0.225	0.196	0.167
Informative	Mean	0.585	1.208	0.807	1.166	0.999	0.796	0.744
	Std. dev.	0.107	0.407	0.169	0.225	0.224	0.247	0.232
Panel B: Recent-IPO and past-industry components								
Recent	Mean	0.497	0.616	0.500	0.516	0.431	0.567	0.599
	Std. dev.	0.459	0.571	0.522	0.419	0.353	0.601	0.439
Past industry	Mean	0.569	0.319	0.540	0.462	0.535	0.467	0.437
	Std. dev.	0.445	0.657	0.496	0.410	0.318	0.572	0.455
Panel C: Pre- vs. post-EU Regulation								
Standard	Pre-reg.	0.981	0.844	0.961	0.857	0.827	0.943	0.951
	Post-reg.	1.017	0.912	0.994	0.969	0.966	0.996	0.993
	Diff.	$+0.036^{*}$	$+0.068$	$+0.033$	$+0.112^{***}$	$+0.138^{***}$	$+0.053^{**}$	$+0.042^{**}$
Informative	Pre-reg.	0.600	1.237	0.805	1.169	0.998	0.763	0.778
	Post-reg.	0.572	1.185	0.809	1.163	0.999	0.824	0.715
	Diff.	-0.028^{**}	-0.052	$+0.004$	-0.006	$+0.001$	$+0.061^{**}$	-0.064^{**}
Recent	Pre-reg.	0.472	0.471	0.484	0.532	0.381	0.498	0.547
	Post-reg.	0.518	0.729	0.513	0.504	0.468	0.624	0.643
	Diff.	$+0.046$	$+0.257^{***}$	$+0.029$	-0.029	$+0.086^{*}$	$+0.126^{*}$	$+0.096^{*}$
Past industry	Pre-reg.	0.618	0.456	0.574	0.422	0.538	0.546	0.502
	Post-reg.	0.532	0.225	0.514	0.491	0.532	0.407	0.386
	Diff.	-0.086	-0.231^{**}	-0.060	$+0.069$	-0.006	-0.139^{**}	-0.116^{**}
N		263	239	261	251	232	263	260

Notes: Decomposition: $\widetilde{\mathbf{w}}_{\text{tot},i} = a_{\text{rec},i}\,\widetilde{\mathbf{w}}_{\text{rec},i} + a_{\text{ind},i}\,\widetilde{\mathbf{w}}_{\text{ind},i} + \boldsymbol{\varepsilon}_i$. Standard content = $\hat{a}_{\text{rec}} + \hat{a}_{\text{ind}}$; Informative content = $\sum_w |\varepsilon_{i,w}|$. Pre-Regulation: IPOs filed before 21 July 2019; Post-Regulation: on or after that date. Differences (Post-reg - Pre-reg)tested with two-sample t-tests. Significance: $^{***}\,p < 0.01$, $^{**}\,p < 0.05$, $^{*}\,p < 0.10$.

5. Empirical Strategy

Section 4 develops the key text-based measures that are the foundation for this section. By comparing each prospectus to relevant peer documents, we decompose disclosure into standard and informative content and show how these evolved around the Regulation. With these measures in place, we now turn to testing whether the Regulation affected disclosure and whether these changes are related to IPO pricing outcomes.

The empirical analysis is structured around the two predictions formulated in Section 4.2. The first step examines whether the Regulation changed the informational composition of IPO prospectuses (Prediction 1). The second step examines whether the balance of standard and informative content within a prospectus is associated with first-day underpricing, and whether that relationship is moderated by the regulatory regime (Prediction 2). Both steps build directly on the content measures constructed in Sections 4.3–4.5.

5.1 Testing Prediction 1 – Content and Regulation

To test whether the Regulation altered the informational composition of prospectuses, we estimate OLS regressions built on those used by Hanley and Hoberg (2010). These regressions have a dependent variable that is either the standard or the informative content scalar for IPO i in section s, as derived from the decomposition in Section 4.5. For each section and content type, we estimate two specifications:

$$\text{Content}_{i,s} = \alpha + \beta_1 \cdot \text{EUReg}_i + \beta_2 \cdot \text{EUGrowth}_i + \gamma \cdot \mathbf{Controls}_i + \delta_t + \delta_{\text{ind}} + \varepsilon_i \quad (7)$$

where EUReg_i equals 1 for IPOs filed after 21 July 2019, and EUGrowth_i equals 1 for IPOs using the EU Growth Prospectus format. Since EU Growth firms are a strict subset of post-Regulation issuers (EUGrowth $\subset$ EUReg), the two dummies are separately identified: EUReg captures the effect of the full Regulation on firms filing a standard prospectus, while EUGrowth captures the incremental effect of the Growth format relative to that baseline. Column (1) includes the regime dummies with year and industry fixed effects but without controls. Column (2) adds the full set of firm and market controls. If coefficients shift materially between columns, this indicates that the apparent regulatory effect is driven by the composition of firms going public under each regime rather than the Regulation itself.

Year fixed effects (δ_t) and industry group fixed effects (δ_{ind}) absorb aggregate time trends and sector-level disclosure norms. Standard errors are clustered two-way by industry group and year throughout, following the approach of Hanley and Hoberg (2010). This accounts for the concentration of IPO activity within years, most notably the 2021 boom (as seen in Appendix G, Table 3 and 4), while capturing residual correlation across firms within the same industry.

We estimate all specifications separately for the total document and each of the six prospectus sections, though the main text focuses on the total document, Summary, and Risk Factors sections, which were subject to significant regulatory changes. Full section-level results are reported in Appendix G.

5.2 Testing Prediction 2 – Content and Pricing

To test whether prospectus content is associated with underpricing and whether that link is moderated by the regulatory regime, we estimate two complementary sets of regressions, both using the initial return (IR) as the primary pricing outcome.[8]

The first set follows a step-by-step structure to assess how sensitive the content coefficients are to the inclusion of controls and regulatory dummies. The dependent variable is winsorised at the 1st and 99th percentiles to mitigate the influence of outliers; robustness to alternative winsorisation levels is examined in Appendix G Table 30. Column (1) regresses IR on standard and informative content with year and industry fixed effects only. Column (2) adds the Regulation and EU Growth dummies. Column (3) adds the full set of firm and market controls:

$$\begin{aligned}\mathrm{IR}_i = &\alpha + \beta_1 \cdot \mathrm{Stand}_{i,s} + \beta_2 \cdot \mathrm{Inf}_{i,s} + \beta_3 \cdot \mathrm{EUReg}_i + \beta_4 \cdot \mathrm{EUGrowth}_i \\ &+ \gamma \cdot \mathbf{Controls}_i + \delta_t + \delta_{\mathrm{ind}} + \varepsilon_i \end{aligned} \tag{8}$$

If the content coefficients shift substantially across columns, this would indicate that regulatory timing or firm characteristics confound the raw content-pricing relationship. If they remain stable, this supports a more direct interpretation.

The second set holds controls fixed across all columns and instead varies the treatment of the regulatory regime. Column (1) is the baseline specification with full controls. Column (2) adds the EU regime dummies additively. Column (3) further introduces four interaction terms to test whether the content-pricing relationship differs across regulatory regimes:

$$\begin{aligned}\mathrm{IR}_i = &\alpha + \beta_1 \cdot \mathrm{Stand}_{i,s} + \beta_2 \cdot \mathrm{Inf}_{i,s} + \beta_3 \cdot \mathrm{EUReg}_i + \beta_4 \cdot \mathrm{EUGrowth}_i \\ &+ \beta_5 \cdot \big(\mathrm{Stand}_{i,s} \times \mathrm{EUReg}_i\big) + \beta_6 \cdot \big(\mathrm{Inf}_{i,s} \times \mathrm{EUReg}_i\big) \\ &+ \beta_7 \cdot \big(\mathrm{Stand}_{i,s} \times \mathrm{EUGrowth}_i\big) + \beta_8 \cdot \big(\mathrm{Inf}_{i,s} \times \mathrm{EUGrowth}_i\big) \\ &+ \gamma \cdot \mathbf{Controls}_i + \delta_t + \delta_{\mathrm{ind}} + \varepsilon_i \end{aligned} \tag{9}$$

The null hypothesis that the EU regime does not moderate the content-pricing channel (H_0: $\beta_5 = \beta_6 = \beta_7 = \beta_8 = 0$) is tested formally using a joint Wald F-test. Rejection would indicate that the Regulation altered not only what issuers disclose, but also how the market prices a given unit of standard or informative content.

Fixed effects, controls, and the two-way clustering structure are identical to those used in the content regressions in Section 5.1. All regressions are estimated separately for the total document, the Summary section, and the Risk Factors section as the main specifications, with full section-level results in Appendix G.

5.3 Limitations with the Methodology

A key limitation is that prospectus content is likely endogenous to expected pricing outcomes. Firms that anticipate difficult market conditions may choose to use more standardized language, meaning that content and pricing are jointly determined

[8] As documented in Section 4.1, 77% of Swedish IPOs are issued at a fixed price within the indicative range, rendering the bookbuilding price revision (ΔP) uninformative for the majority of the sample; results for ΔP and $|\Delta P|$ are reported in Appendix G.

rather than clearly causal. Addressing this would require an instrument that affects content without directly influencing pricing, but no such variable is available in our data. The results should therefore be interpreted as associations rather than causal effects.

Another limitation relates to data quality. Unlike the US, Sweden does not provide public filings in HTML format, but only in PDF. The prospectuses are converted from PDF to HTML to machine-readable text, which can introduce noise through formatting issues, incorrect section splits, or encoding errors – especially for older documents. Although we have manual checks, some degree of measurement error is likely in the word vectors. This is especially important for the section-level analysis, since words may sometimes be assigned to the wrong section. Smaller sections, such as Background & Motive, are more sensitive to this type of error because a few misplaced words can have a larger effect on the section's measured content.

Thirdly, with 269 IPOs, our sample is substantially smaller than the 1,750 IPOs used by Hanley and Hoberg (2010),[9] which reduces statistical power and limits the precision of the standard content estimation, particularly for sections with relatively few words, such as Background & Motive.

6. Results

6.1 Results – Prediction 1: Content vs. Regulation

Having established the sources of content similarity in Table 2, we now turn to the first core question of our analysis: whether the Regulation affected the informational composition of prospectuses. Table 4 tests whether the Regulation changed standard and informative content in the Total Document, Summary, and Risk Factors sections (remaining sections in Appendix G). The key independent variables are EU Regulation and EU Growth.

A first point to note from Table 4 is that the EU Regulation coefficient is consistently insignificant across all sections and for both dependent variables. This is an important null result. Once firm characteristics and fixed effects are included, the Regulation itself does not seem to have affected the composition of content.[10] This suggests that the increase in standardization after 2019 is at least partly driven by the types of firms going public in that period, rather than the Regulation itself.

In contrast, the EU Growth Prospectus variable shows strong and significant effects across several sections. For example, it is associated with increases of around $+0.119^{***}$ in total standard content and $+0.123^{***}$ in total informative content (without controls). This indicates that both informative and standard content increased, meaning the EU Growth format did not increase firm-specific disclosure at the expense of standardized language. In other words, the EU Growth format may have made the prospectus more structured while simultaneously allowing issuers to provide more firm-specific information. One interpretation is that the reduced disclo-

[9] Appendix G, Figure 5 plots power curves for the IR specifications, showing that detecting effects observed here at significance levels would require a sample of approximately 800–1,000 IPOs.

[10] This does not contradict the findings in Table 3 Panel C, where standardization increases after 2019. The two results are not directly comparable: Table 3 uses raw averages, while Table 4 controls for firm differences.

sure burden under the Growth regime allowed smaller firms to shift focus away from compliance-driven standardization and instead emphasize more firm-specific information – though this effect is firm-type dependent rather than the broad regulatory shift anticipated under Prediction 1.

Another relevant result, visible in Table 4, is the role of filing size. Log Filing Size is significantly negative for standard content in multiple sections (e.g. -0.043^{***} for the total document and -0.051^{***} for the Risk Factors section). This implies that larger offerings tend to rely less on standardized language and produce more informative disclosures. An interpretation is that larger IPOs, which typically have more resources, invest more in firm-specific content.

Table 4: Determinants of Standard and Informative Content (Selected Sections)

	Dep. var.: Standard content						*Dep. var.: Informative content*					
	Tot		*Sum*		*Risk*		*Tot*		*Sum*		*Risk*	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
EU Regulation	−0.010	0.039	−0.472	−0.285	−0.079	0.029	0.041	0.049	−0.077	0.029	0.090	0.130
	(−0.16)	(1.06)	(−1.53)	(−0.68)	(−0.86)	(0.36)	(1.46)	(1.32)	(−0.27)	(0.08)	(1.03)	(1.25)
EU Growth Prospectus	0.119***	0.002	−0.005	−0.075**	0.182***	0.031	0.123***	0.095***	0.129	0.063	0.136***	0.070**
	(2.99)	(0.05)	(−0.10)	(−2.26)	(4.54)	(0.66)	(7.29)	(5.61)	(1.21)	(0.55)	(4.99)	(2.01)
Log Firm Age	—	−0.003	—	0.066***	—	0.004	—	−0.011**	—	0.051**	—	−0.004
		(−0.26)		(2.67)		(0.30)		(−2.06)		(2.48)		(−0.44)
Log Filing Size	—	−0.043***	—	−0.038	—	−0.051***	—	−0.009	—	−0.022	—	−0.027***
		(−5.30)		(−0.94)		(−2.95)		(−1.27)		(−0.71)		(−3.54)
Control variables	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	263	257	239	236	261	255	263	257	239	236	261	255
R^2	0.324	0.571	0.169	0.212	0.324	0.497	0.535	0.575	0.147	0.181	0.370	0.411

Notes: OLS regressions (Equation 7) with two-way industry × year clustered standard errors. Standard content = $\hat{a}_{\text{rec}} + \hat{a}_{\text{ind}}$; Informative = $\sum_w |\varepsilon_{i,w}|$. Additional controls (Column (2)): tech dummy, PE dummy, regulated market dummy, lead underwriter market share, OMX30 30-day return, log # IPOs (180d). See Appendix G Tables 15 and 16 for full results across all seven sections. Pricing variables are winsorised at the 1st and 99th percentiles (baseline). t-statistics in parentheses. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

6.2 Results – Prediction 2: Effects on Price

Using the step-by-step method outlined in Section 5.2, Table 5 presents OLS regressions of the initial return (IR) on: (1) standard and informative content; (2) the EU Regulation and EU Growth dummies; and (3) the full set of firm and market controls.

Across all three selected sections in Table 5, neither standard nor informative content is statistically significant. Adding the EU Regulation and EU Growth dummies in Column (2), does not materially change the coefficients, and adding the full set of controls in Column (3) does not alter the main conclusion. This suggests that prospectus content does not have a clear unconditional relationship with IPO underpricing in the Swedish setting.

Figure 1 confirms this result visually. The confidence intervals for both standard and informative content are wide and consistently include zero across specifications. The direction of the coefficient signs for the Total Document and Risk Factors sections are consistent with Hanley and Hoberg (2010), with standard content generally being positively and informative content negatively, but the estimates are too imprecise to draw a conclusion. Thus, unlike Hanley and Hoberg (2010), we do not find evidence that informative content is associated with lower underpricing in Sweden.

Taken together, Table 5 and Figure 1 point to the same conclusion: the estimates are too imprecise to say whether the relationship between disclosure content and underpricing is positive or negative. While the estimates for the total document and Risk Factors sections show standard content leaning positive and informative content leaning negative – directions consistent with Hanley and Hoberg (2010) – the confidence intervals are wide enough that we can not tell the directions of the findings. This is a difference from Hanley and Hoberg (2010), who find a clear and significant relationship in their US sample: informative content reduces underpricing by approximately 8%, while standard content is associated with larger first-day returns. We find the same directional pattern, but the effect is far weaker and indistinguishable from zero.

This difference is important in connection to the discussion in Section 2.1. Hanley and Hoberg's (2010) framework, informative content reduces underpricing because it substitutes for costly information production during bookbuilding, consistent with the logic of Benveniste and Spindt (1989). If this mechanism were visible in Sweden, informative content should be more clearly associated with pricing accuracy, which is not the case. One potential reason is that 77% of Swedish IPOs in our sample have zero price adjustment, indicating that fixed-price offerings are common. When the offer price is not revised in response to investor demand, the bookbuilding channel through which disclosure affects pricing is partly muted. In such a setting, underpricing may instead function more as a participation discount for uninformed investors, as in Rock (1986)), than as compensation for information revelation during bookbuilding. Price discovery may therefore rely less on prospectus text alone and more on institutional features of the IPO process, including repeated interactions between underwriters and institutional investors, as emphasized by Cornelli and Goldreich (2001). This interpretation helps explain why the content-pricing relationship is weaker than in the US setting studied by Hanley and Hoberg (2010).

Building on the step-by-step structure of Table 5, Table 6 tests whether the pricing role of disclosure differs across regulatory regimes. It does so by replacing the standalone EU Regulation and EU Growth dummies with interaction terms

Table 5: Content and Initial Returns (IR) — Step-by-Step Specifications (Selected Sections)

	Dep. var.: IR (first-day return)								
	Tot			*Sum*			*Risk*		
	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
Standard content	0.117	0.125	0.226	0.010	0.005	−0.002	0.134	0.131	0.156
	(0.77)	(0.69)	(0.84)	(0.14)	(0.07)	(−0.02)	(1.61)	(1.48)	(1.08)
Informative content	0.187	0.280	0.221	0.003	0.008	−0.002	−0.123	−0.104	−0.145
	(0.79)	(0.88)	(0.82)	(0.04)	(0.11)	(−0.03)	(−1.03)	(−0.79)	(−1.21)
EU Regulation	—	−0.136	−0.132	—	−0.082	−0.035	—	−0.106	−0.098
		(−0.95)	(−0.84)		(−0.49)	(−0.19)		(−0.70)	(−0.57)
EU Growth (incr.)	—	−0.050	−0.030	—	−0.012	−0.025	—	−0.011	−0.002
		(−0.38)	(−0.28)		(−0.12)	(−0.23)		(−0.10)	(−0.02)
Control variables	No	No	Yes	No	No	Yes	No	No	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	257	257	251	233	233	230	255	255	249
R^2	0.139	0.144	0.186	0.156	0.157	0.187	0.130	0.132	0.175

Notes: OLS regressions (Equation 8) with two-way industry × year clustered standard errors. Column (1): content and fixed effects only. Column (2): adds EU Regulation and EU Growth Prospectus dummies. Column (3): adds full controls (log firm age, log filing size, tech dummy, PE dummy, regulated market dummy, lead underwriter market share, OMX30 30-day return, log # IPOs (180d)). EU Regulation = 1 for IPOs filed on or after 21 July 2019; EU Growth (incr.) denotes the incremental effect of the EU Growth sub-regime. No content coefficient is statistically significant at conventional levels. See Appendix G Table 17 for full results across all seven sections and all control variables. Pricing variables are winsorised at the 1st and 99th percentiles (baseline). t-statistics in parentheses. Significance: $^{***}\,p < 0.01$, $^{**}\,p < 0.05$, $^{*}\,p < 0.10$.

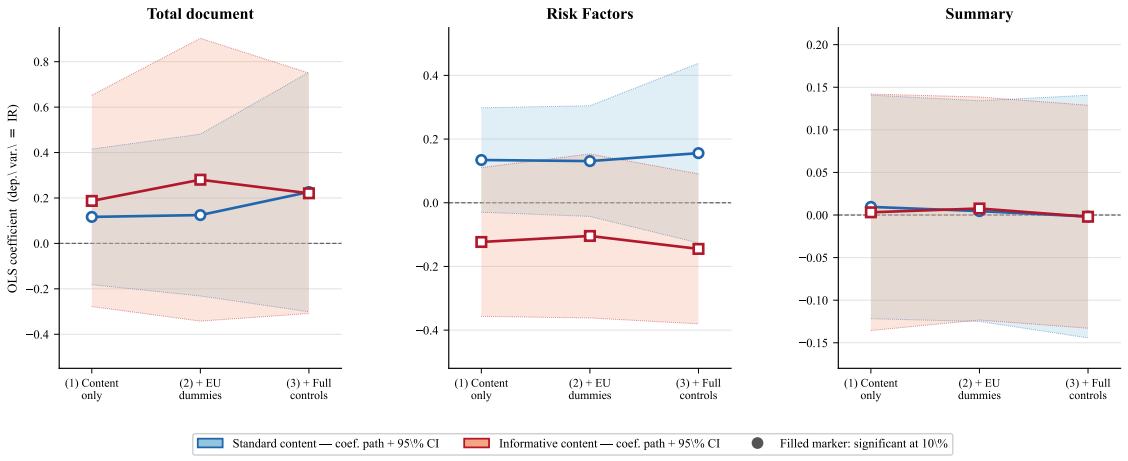


Figure 1: Stability of content coefficients across step-by-step specifications (IR).

Notes: OLS regressions with two-way industry × year clustered standard errors. Dependent variable is IR (first-day return). (1): content and fixed effects only. (2): adds EU Regulation and EU Growth dummies. (3): adds firm-level and market controls (see Table 5). Shaded bands denote 95% confidence intervals. No coefficient is statistically significant at conventional levels across any specification or section.

between these regulatory variables and the content measures.

Importantly, once these interaction terms are included, the standalone EU Regulation and EU Growth coefficients should not be interpreted as average regime effects. Instead, they capture the effect of the regulatory regime when the content variables equal zero, which is not economically meaningful in this setting. The interpretation therefore focuses on the interaction terms and the joint Wald F-test reported below, rather than on the standalone regime coefficients.

The most notable shift in Table 6 appears in the Risk Factors section. Standard content, is significant in Panel C ($\hat{\beta} = 0.251$, $t = 3.11$, $p < 0.01$), while EU Regulation increases considerably once the interaction terms are included ($\hat{\beta} = 0.521$, $t = 2.33$, $p < 0.05$), with a similar pattern in the Total Document section (0.557^{**}). The R^2 also rises from approximately 0.173–0.187 in Panel A to 0.208–0.227 in Panel C, with the largest gain precisely in Risk Factors. This suggests that the interaction terms add explanatory power and that the relationship between standard risk language and first-day returns differs across regulatory regimes.

Because the individual interaction coefficients can be difficult to interpret in isolation, we assess the overall moderation effect through a joint Wald F-test. The test evaluates whether all four interaction terms are jointly equal to zero and is reported in Table 7.

Table 6: Content and Initial Return (IR) — Selected Sections

	Dep. var.: IR (first-day return)								
	Tot			*Sum*			*Risk*		
	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
Standard content	0.233	0.226	0.301*	0.001	−0.002	0.015	0.157	0.156	0.251***
	(0.82)	(0.84)	(1.67)	(0.02)	(−0.02)	(0.17)	(1.11)	(1.08)	(3.11)
Informative content	0.163	0.221	0.221	−0.005	−0.002	−0.002	−0.158	−0.145	−0.021
	(0.63)	(0.82)	(0.74)	(−0.08)	(−0.03)	(−0.02)	(−1.38)	(−1.21)	(−0.15)
EU Regulation	—	−0.132	0.557**	—	−0.035	0.212	—	−0.098	0.521**
		(−0.84)	(2.12)		(−0.19)	(0.84)		(−0.57)	(2.33)
EU Growth (incr.)	—	−0.030	−1.358**	—	−0.025	−0.417**	—	−0.002	−0.535
		(−0.28)	(−2.13)		(−0.23)	(−2.14)		(−0.02)	(−1.40)
Stand × EU Reg.	—	—	−0.469*	—	—	−0.111	—	—	−0.537***
			(−1.79)			(−0.96)			(−2.74)
Inf × EU Reg.	—	—	−0.367	—	—	−0.119	—	—	−0.143
			(−0.76)			(−0.79)			(−0.82)
Stand × EU Growth	—	—	0.669	—	—	0.140	—	—	1.100***
			(0.67)			(0.53)			(2.99)
Inf × EU Growth	—	—	1.051	—	—	0.212	—	—	−0.711**
			(0.90)			(1.01)			(−1.99)
Control variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	251	251	251	230	230	230	249	249	249
R^2	0.183	0.186	0.209	0.187	0.187	0.208	0.173	0.175	0.227

Notes: OLS regressions (Equation 9) with two-way industry × year clustered standard errors. Dependent variable is IR (first-day return). Stand. = Standard Content. Inf. = Informative Content. Column (1): Panel A (content and controls only). Column (2): Panel B (adds EU Regulation and EU Growth dummies additively). Column (3): Panel C (further adds four content × EU regime interaction terms). EU Regulation = 1 after 21 July 2019; EU Growth ⊂ EU Regulation. Controls (log firm age, log filing size, tech dummy, PE dummy, regulated market dummy, lead underwriter market share, OMX30 30-day return, log # IPOs (180d)) included in all columns. See Appendix G Tables 18, 19, and 20 for full results across all seven sections. Pricing variables are winsorised at the 1st and 99th percentiles (baseline). t-statistics in parentheses. Significance: $^{***}\,p < 0.01$, $^{**}\,p < 0.05$, $^{*}\,p < 0.10$.

Joint Wald F-test

Table 7 reports the joint Wald F-test of whether the Regulation moderates the content-pricing relationship, testing the null hypothesis that all four interaction coefficients are jointly zero. As the step-by-step regressions in Tables 5 and 6 consistently fail to find a significant unconditional relationship between either content type and initial return, the result of Table 7 aims relate directly to Prediction 2: looking at the moderation effect of the Regulation on the content-pricing channel for selected sections (Total Document, Risk Factors, and Summary).

Table 7: Joint Wald F-Test: EU Regime Moderation of Content–Pricing Channel

	IR (first-day return)		
Section	F	df_2	p
Tot	5.319***	194	0.0004
Sum	3.535***	177	0.0084
Risk	3.621***	193	0.0072

Notes: Joint Wald F-test of H_0: $\beta_{\text{Stand}\times\text{EUreg}} = \beta_{\text{Inf}\times\text{EUreg}} = \beta_{\text{Stand}\times\text{EUgro}} = \beta_{\text{Inf}\times\text{EUgro}} = 0$, i.e. the EU regime does not moderate the content–pricing relationship. F-statistics from Panel C regressions (content × EU interaction, full controls, year and industry FE). Degrees of freedom: $df_1 = 4$ restrictions (constant across rows); $df_2 = N - k$ reported per section. p-values use the $F(4, df_2)$ distribution. Two-way industry × year clustered standard errors. See Appendix G Table 21 for results across all seven sections. Significance: $^{***}\,p < 0.01$, $^{**}\,p < 0.05$, $^{*}\,p < 0.10$.

In the Wald F-test, Table 7, the null is rejected for initial returns across all three main sections – total document ($F = 5.319$, $p = 0.0004$), Summary ($F = 3.535$, $p = 0.0084$), and Risk Factors ($F = 3.621$, $p = 0.0072$) – at the 1% significance level. The rejection across all three sections is internally consistent, and the F-statistics between 3.5 and 5.3 suggest that no single section drives the moderation. This means that the relationship between content and pricing is not uniform across the sample, but systematically different before and after the Regulation. The content-pricing link that Prediction 2 anticipated exists, but is conditional on the Regulation rather than fully unconditional. The standard content coefficient in Risk Factors from Table 6 becoming significant ($\hat{\beta} = 0.251$, $p < 0.01$) only when the interactions are included is a concrete illustration of this: the Regulation did not only change what issuers must disclose, but also how the market prices a given unit of disclosure.

However, the same is not true for the Summary section, for although the Wald F-test rejects for IR ($p = 0.0085$), the underlying point estimates in Tables 5 and 6 show that all Summary content coefficients are indistinguishable from zero throughout. The joint test can reject even when no individual interaction term is large, so the rejection here is more statistical than economic.

On a broader point – the Regulation reshaped the information environment for Swedish IPOs – the strongest empirical support is found neither in the content composition regressions (Prediction 1) nor in the pricing regressions (Prediction 2), but in the interaction tests. The key contribution is therefore slightly different from the original predictions: the Regulation did not mechanically increase firm-specific disclosure, and it did not produce a simple reduction in underpricing through increasing informative content. Instead, the Regulation appears to have altered how the market prices disclosure content. This connects to the mechanism documented by Chang, Ljungqvist, and Tseng (2023), where the introduction of EDGAR affected analyst

behavior not by increasing the amount of information available, but by changing how mandatory disclosure was processed and interpreted. In a similar way, our results suggest that the Regulation changed investors' interpretation of prospectus language, not just its structure. A possible explanation is that, by forcing the Risk Factors section to be more specific, material, and ranked in order of importance, boilerplate language became easier to identify and therefore more informative about a firm's underlying state. Post-Regulation, a unit of standard risk content may therefore signal more information asymmetry than pre-Regulation, explaining why the content-pricing relationship strengthened.

6.3 Robustness Checks

The findings in Sections 6.1 and 6.2 depend on several empirical choices: how the joint Wald F-test holds across alternative specifications, how standard errors are clustered, how the sample is split around the Regulation and whether a Chow test confirms a structural break, how the price-revision channel is treated given Sweden's prevalence of fixed-price offers, how underwriter fixed effects are handled, how the decomposition window is defined, how outliers are winsorised, whether EU Growth firms are confounded with multilateral trading facility listings, and how the 2021 IPO boom is handled in inference. We test each of these choices in turn to establish which conclusions from Sections 6.1 and 6.2 are robust to alternative specifications and which are sensitive to specific modelling choices.

PCA Dimensionality Reduction of Word Vectors

The word vectors used in the baseline analysis span a 3,496-dimensional vocabulary space (constructed as described in Section 4.3), far exceeding the 269 observations in our sample. While the Hanley and Hoberg (2010) methodology does not require the word vectors to be full rank – the decomposition operates via OLS on peer-average vectors, not on the raw vocabulary matrix – a legitimate concern is that many vocabulary dimensions are redundant and that the standard and informative content measures are sensitive to noise in the high-dimensional space. We address this by re-running the full analysis[11] after compressing the word vectors via Principal Component Analysis (PCA); for the methodology, see Appendix F.

There is a substantial increase in the informative content mean as more PCA components are retained, rising from 3.82 at $k = 40$ to 6.56 at $k = 125$, remaining well above the baseline of 0.58 throughout.[12] This is, however, a mechanical artifact rather than a substantive finding. PCA discards the low-variance vocabulary tail, and the subsequent L^2-normalization re-scales the residual space, meaning that the OLS residuals from the decomposition land in a geometrically different subspace, inflating the sum of absolute residuals that constitute the informative content measure. The cross-sectional rank ordering of prospectuses is preserved even though absolute levels are not, which is what matters for our regressions. A related observation is that the mean standard content falls slightly while its standard deviation increases substantially, indicating that PCA increases cross-sectional variation in the

[11] The PCA robustness re-runs all content and pricing regressions from Sections 6.1 and 6.2 using standard and informative content measures derived from PCA-compressed word vectors rather than the full 3,496-dimensional vocabulary space. The number of components retained is determined by the elbow criterion applied to the explained variance curve, as described in Appendix F.

[12] See Table 22 for full results.

coefficients. Importantly, this greater dispersion gives the regressions more statistical power to detect pricing effects, making the persistent insignificance in Panel A (Table 9) a stronger confirmation of the null result than the baseline alone.

The baseline's significant EU Growth effect disappears entirely under all three PCA specifications, as seen in Table 8. The EU Growth effect on informative content is therefore not robust to dimensionality reduction. This is most likely explained by correlated vocabulary patterns in the high-variance dimensions discarded by PCA: the signal appears to be partly driven by words concentrated in the low-variance tail of the distribution of word roots – words specific to EU-era prospectuses – rather than a genuine shift in the structure of disclosed content. Specifically, the finding may reflect a terminology change in prospectus language under the new regulatory template rather than a real increase in unique, firm-specific information. By contrast, the EU Regulation coefficient on both content types is consistently small and insignificant across all four specifications, which strengthens our earlier conclusion that the shift to the Regulation did not systematically alter the level of either standard or informative content.

Table 8: PCA Robustness — EU Regulation Effects on Content

	Baseline 3,496 dims	PCA-40 $k = 40$	PCA-80 $k = 80$	PCA-125 $k = 125$
Standard content (z)				
EU Regulation	+0.2667	+0.0881	−0.0904	−0.0927
	(+1.03)	(+0.33)	(−0.37)	(−0.39)
EU Growth	+0.0064	+0.1166	+0.0363	+0.0202
	(+0.03)	(+0.52)	(+0.16)	(+0.09)
N	257	257	257	257
Informative content (z)				
EU Regulation	+0.4328	+0.3012	+0.3822	+0.1084
	(+1.45)	(+0.54)	(+0.78)	(+0.21)
EU Growth	+0.8909***	−0.1463	−0.0279	−0.0197
	(+5.07)	(−0.61)	(−0.11)	(−0.08)
N	257	257	257	257

Notes: OLS regressions (Equation 7) of standardized (z-scored) content variables on EU regulation dummies, estimated on PCA-compressed word vectors as an alternative to the baseline specification in Table 4. Baseline uses the full 3,496-dimensional word vectors (total-document specification); PCA-k replaces these with the first k principal components, retaining 80%, 90%, and 95% of total variance for $k = 40$, 80, and 125 respectively. Full controls: log firm age, log filing size, tech dummy, PE dummy, regulated market dummy, lead underwriter market share, OMX30 30-day return, log # IPOs (180d). Year and industry fixed effects included. Two-way industry × year cluster-robust standard errors. t-statistics in parentheses. Significance: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Table 9: PCA Robustness — Content and Initial Return (IR)

	Panel A: Content + controls				*Panel C: EU interaction*			
	Baseline 3,496 dims (1)	PCA-40 $k = 40$ (2)	PCA-80 $k = 80$ (3)	PCA-125 $k = 125$ (4)	Baseline 3,496 dims (5)	PCA-40 $k = 40$ (6)	PCA-80 $k = 80$ (7)	PCA-125 $k = 125$ (8)
Standard (z)	+0.0333	+0.0093	+0.0032	+0.0017	—	—	—	—
	(+0.63)	(+0.56)	(+0.19)	(+0.10)				
Informative (z)	+0.0173	+0.0468**	+0.0281	+0.0208	—	—	—	—
	(+0.66)	(+2.24)	(+1.23)	(+0.91)				
Stand (z) × EU Reg.	—	—	—	—	−0.0685	+0.0061	+0.0188	+0.0260
					(−1.46)	(+0.18)	(+0.50)	(+0.64)
Inf (z) × EU Reg.	—	—	—	—	−0.0499	−0.0058	+0.0155	+0.0209
					(−0.93)	(−0.14)	(+0.36)	(+0.45)
Stand (z) × EU Growth	—	—	—	—	+0.0912	+0.0754*	+0.0703*	+0.0686*
					(+0.69)	(+1.72)	(+1.67)	(+1.67)
Inf (z) × EU Growth	—	—	—	—	+0.1402	+0.0194	−0.0328	−0.0434
					(+0.98)	(+0.22)	(−0.33)	(−0.42)
Control variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	251	251	251	251	251	251	251	251
R^2	0.175	0.183	0.172	0.170	0.202	0.193	0.186	0.184

Notes: OLS regressions (Equation 9) with two-way industry × year clustered standard errors. Dependent variable is initial return (IR). Panel A: content and firm controls with year and industry fixed effects. Panel C adds four content × EU regime interaction terms. Content variables (Standard, Informative) are z-scored within each specification so that coefficients are comparable across compression levels. Baseline uses the full 3,496-dimensional word vectors; PCA-k compresses to k principal components before the Hanley and Hoberg (2010) decomposition, retaining 80%, 90%, and 95% of total variance for $k = 40$, 80, and 125 respectively. Controls: log firm age, log filing size, tech dummy, PE dummy, regulated market dummy, lead underwriter market share, OMX30 30-day return, log # IPOs (180d). For PCA robustness results on price adjustment (ΔP) and absolute price adjustment ($|\Delta P|$), see Appendix G Table 23. t-statistics in parentheses. Significance: $^{***}\,p < 0.01$, $^{**}\,p < 0.05$, $^{*}\,p < 0.10$.

In Panel A of Table 9, the coefficients on Standard(z) and Informative(z) are small and consistently positive across all models, but none are statistically significant. The only exception is Informative(z) at PCA-40 ($t = 2.24$), which does not persist at higher k values and is likely a statistical outlier. Overall, the relationship between content and pricing for IR appears genuinely weak and not driven by the high-dimensional vocabulary space, supporting our main findings.

In Panel C, nearly all interaction coefficients are insignificant across specifications. The one partial exception is the Stand(z) × EU Growth coefficient, which is marginally significant at PCA-80 and PCA-125 ($p < 0.10$) but insignificant in the baseline. Rather than contradicting the main analysis, this could indicate that, once vocabulary noise is removed, we can detect a slight latent pattern of standard content being priced more positively for EU Growth prospectuses. However, given that this is a single interaction term significant only at the 10% level and absent in the baseline, the evidence is weak and inconclusive.

Overall, the PCA robustness check supports the findings in Sections 6.1 and 6.2: the content-pricing relationship for IR is weak regardless of how the vocabulary space is compressed, and the Regulation level effects are stable throughout. The principal exception is the EU Growth effect on informative content from Table 6, which is sensitive to the dimensionality of the word vector space and should therefore be interpreted with caution.

Standard Error Specification

Following Hanley and Hoberg (2010), we cluster standard errors two-way by industry and year. This approach captures the concentrated 2021 IPO boom (44 listings in Q2 alone, Appendix G) within the year cluster, while still accounting for residual correlation across firms within the same industry. To ensure that this specification does not drive the significance patterns reported in Sections 6.1 and 6.2, we re-estimate all key results using three alternative covariance structures: HC1 (no clustering), one-way clustering by year, and three-way clustering by industry, year, and underwriter.[13] The check covers six cells; definitions are provided in Appendix C.

Table 10 confirms that the results from Sections 6.1 and 6.2 are not driven by the choice of covariance structure. No coefficient crosses conventional significance thresholds under any of the four specifications, and no sign changes occur across cells C1–C5. The t-statistics are largely stable across HC1, year-clustered, and two-way specifications, except for standard content in C1 and C2, where year-only clustering yields slightly higher precision ($t = +1.62$ versus the baseline $t = +0.77$) – most likely a consequence of having few year clusters rather than a genuine strengthening of the effect. The three-way industry × year × underwriter specification produces t-statistics identical to the two-way baseline throughout, as the PSD-fallback is triggered in every case, given the thin cell structure of the sample. For C6, the EU Regulation coefficient remains small and insignificant under all four specifications (t ranging from +0.06 to +2.11 across standard error structures, with the significant value arising only under year-only clustering for the same mechanical reason as C1–C2), while the EU Growth coefficient retains its positive sign and approximate magnitude throughout, confirming that neither the null on Regulation nor the EU Growth effect is driven by the choice of standard error estimator. The inference

[13] When the three-way covariance matrix is not positive semi-definite, we revert to the two-way estimator.

reported in the main analysis is therefore robust to the full range of error structures considered.

Table 10: Robustness Check — Standard-Error Specification

Cell	Variable	Section	HC1	Year	Ind×Year	Ind×Yr×UW
C1	Informative	Tot	+0.187 (+0.80)	+0.187 (+0.91)	+0.187 (+0.79)	+0.187 (+0.79)
C1	Standard	Tot	+0.117 (+0.72)	+0.117 (+1.62)	+0.117 (+0.77)	+0.117 (+0.77)
C2	Informative	Tot	+0.221 (+0.92)	+0.221 (+1.06)	+0.221 (+0.82)	+0.221 (+0.82)
C2	Standard	Tot	+0.226 (+0.74)	+0.226 (+1.43)	+0.226 (+0.84)	+0.226 (+0.84)
C3	Informative	Risk	−0.145 (−1.00)	−0.145 (−0.73)	−0.145 (−1.21)	−0.145 (−1.21)
C3	Standard	Risk	+0.156 (+1.42)	+0.156 (+1.60)	+0.156 (+1.08)	+0.156 (+1.08)
C6	EU Growth	Tot	+0.002 (+0.06)	+0.002 (+0.10)	+0.002 (+0.05)	+0.002 (+0.05)
C6	EU Reg.	Tot	+0.039 (+0.76)	+0.039** (+2.11)	+0.039 (+1.06)	+0.039 (+1.06)

Notes: OLS coefficients with t-statistics in parentheses under four cluster-robust covariance specifications: HC1 (heteroskedasticity-robust, no clustering), one-way year clustering, two-way industry × year clustering (baseline), and three-way industry × year × underwriter clustering. Cells C1–C6 are defined in Appendix C: C1 re-estimates Total IR at step (1) (content and fixed effects only); C2 re-estimates Total IR at step (3) (full controls); C3 re-estimates Risk IR at step (3); C6 stress-tests the EU Regulation and EU Growth coefficients on total standard and informative content (Table 4). Full results including ΔP specifications (C4–C5) and all sections are reported in Appendix G Table 24. The joint Wald F-test of EU-regime moderation is stress-tested separately in Table 11. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Joint Wald F-test Robustness

Because the joint Wald F-test in Table 11 is the central result of Section 6.2, we stress-test the F-statistic itself rather than individual coefficients. An incorrect rejection of H_0 driven by the choice of covariance structure, the 2021 boom year, or the difference between regulated and MTF venues could invalidate the central conclusion without any individual coefficient appearing unstable.

Table 11: Wald F-Robustness — Joint EU-Regime Moderation Across Batteries

Tests	Section	F	df_1	df_2	p	N
Baseline (= Table 7)	Total	5.319	4	194	0.0004***	251
	Risk	3.621	4	193	0.0072***	249
Cluster = HC1	Total	1.112	4	194	0.3520	251
	Risk	3.599	4	193	0.0074***	249
Cluster = Year	Total	11.189	4	194	0.0000***	251
	Risk	6.434	4	193	0.0001***	249
Excl. 2021	Total	2.211	4	124	0.0716*	179
	Summary	4.330	4	110	0.0027***	161
	Risk	0.867	4	123	0.4858	177

Notes: Joint Wald F-test of H_0: all four content × EU regime interaction coefficients equal zero. The Baseline row replicates Table 7. HC1 denotes heteroskedasticity-robust standard errors without clustering; Year denotes one-way year clustering. Excl. 2021 drops the IPO boom year; Summary is reported for this test as it crosses significance in that sub-sample. Full results are reported in Appendix G Table 25. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

The Wald F-test rejects the null of no Regulation moderation at the 1% level across all three sections in the baseline specification, validating the Panel C results of

Section 6.2. One notable exception arises under HC1 clustering for the Total Document, where the F-statistic drops from 5.319 to 1.112 and loses significance entirely. This is consistent with the sample composition results below, which show that the 2021 IPO boom meaningfully affects the total document finding: HC1 treats each observation as independent and therefore cannot account for the strong within-year residual correlation generated by the boom, understating the standard errors and inflating the F-statistic. This validates our choice of a covariance structure that clusters at least by year. Excluding 2021 entirely, the Total Document Wald F becomes only marginally significant and the Risk Factors section loses significance, while the Summary section remains strongly significant. Taken together, this suggests that the boom year amplifies our moderation finding in certain sections but is not its sole driver.

Sub-period Stability

The interpretation in Section 6.2 of the Regulation as moderating the content-pricing channel is implemented via the interaction terms between content and the Regulation variable in Panel C. To complement this parametric approach, we conduct a non-parametric test by splitting the sample at 21 July 2019, the effective date of the Regulation. We then re-estimate Total IR (with full controls), Risk IR (with full controls), and Total ΔP (using the Panel A baseline specification) separately for each sub-period. Finally, we report a Chow-style F-test assessing the joint equality of the content coefficients across the two periods.

Table 12: Robustness Check — Pre/Post-EU Regulation Sub-Period Split

Specification	Variable	Period	Coef.	t	N
Total IR (full)	Standard	All	+0.226	+0.84	251
		Pre	+0.088	+0.35	113
		Post	+0.276	+0.40	138
	Informative	All	+0.221	+0.82	251
		Pre	+0.478**	+2.27	113
		Post	+0.086	+0.17	138
Risk IR (full)	Standard	All	+0.156	+1.08	249
		Pre	+0.278**	+2.36	111
		Post	+0.097	+0.35	138
	Informative	All	−0.145	−1.21	249
		Pre	−0.101	−0.52	111
		Post	−0.325	−1.46	138
Chow-style F-test (H_0: content coefficients identical pre/post):					
Total IR (full)	—	—	$F = 0.144$	$df = (2, 195)$	$p = 0.8656$
Risk IR (full)	—	—	$F = 0.830$	$df = (2, 194)$	$p = 0.4375$

Notes: OLS regressions with full controls and year and industry fixed effects; two-way industry × year clustered standard errors. 'Pre' = filings before 21 July 2019; 'Post' = on or after. The Chow-style F-test pools both periods with content × period interactions and tests their joint equality to zero. Full results including ΔP specifications are reported in Appendix G Table 26. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

None of the Chow F-tests in Table 12 reject the null hypothesis that content coefficients are identical across the pre- and post-Regulation periods. This is expected given the inherent power limitation – splitting the sample at 21 July 2019 leaves roughly half the observations in each sub-period, substantially reducing the test's

ability to detect structural breaks. Notably, the pre-period coefficient on informative content in Total IR is significant ($+0.478^{**}$, $t = 2.27$), while the post-period estimate falls to $+0.086$ and loses significance entirely. This pattern is consistent with a moderating effect: the content-pricing relationship was stronger before the Regulation and weakened thereafter, further supporting our conclusion that the Regulation altered how the market prices prospectus content.

Bookbuilding Sub-sample and the $\Delta P = 0$ Finding

77% of the non-missing price adjustment observations in our sample are exactly zero, reflecting the prevalence of fixed-price offers in Sweden – a setting that Hanley and Hoberg's (2010) bookbuilding mechanism does not directly address. To verify that this institutional feature does not distort our inferences, we re-estimate every ΔP and $|\Delta P|$ regression on the bookbuilding subsample ($\Delta P \neq 0$).

Table 13: Robustness Check — Bookbuilding Sub-Sample ($\Delta P \neq 0$)

Section	Panel	Variable	Full sample ($n \approx 255$)	BB only ($n \approx 49$)
Key flip: baseline ΔP results distorted by fixed-price offers				
Risk	A	Informative	-0.191^{***}	$+0.002$
Risk	A	Standard	-0.088	-0.917
Tot	A	Informative	$+0.092$	$+0.954$
Tot	A	Standard	-0.231	-4.150
Moderation finding (Panel C) mostly stable across samples				
Risk	C	Stand × EU Reg.	-0.204	-6.045
Risk	C	Inf × EU Reg.	$+0.071$	$+8.390^{*}$
Tot	C	Stand × EU Reg.	-0.226	-2.774
Tot	C	Inf × EU Reg.	$+0.365$	$+6.876$

Notes: OLS coefficients from ΔP regressions on the full sample and the bookbuilding sub-sample ($\Delta P \neq 0$), with full controls and two-way industry × year clustered standard errors. 77% of ΔP observations are exactly zero (fixed-price offers), which mechanically drives the full-sample ΔP results. The flip in the Risk informative content coefficient (from -0.191^{***} to $+0.002$) confirms that the full-sample ΔP finding is an institutional artifact rather than an informational signal, supporting the decision to lead the analysis with IR. Panel C interaction terms are mostly stable across samples, suggesting the EU-regime moderation finding is not driven by the prevalence of fixed-price offers. The bookbuilding sub-sample is small ($N \approx 49$); stable estimates should be treated as supporting rather than conclusive. Full results in Appendix G Table 27. Significance: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

The informative content coefficient in Panel A of Table 13, flips sign for the Risk Factors section when restricting to the bookbuilding subsample (from -0.191^{***} to $+0.002$), illustrating how the full-sample ΔP results are distorted by the 77% of observations with zero-price revision. The instability of these coefficients once fixed-price IPOs are removed confirms our decision to centre the analysis on IR rather than ΔP in Sections 6.1 and 6.2. The Panel C interaction terms are, however, mostly stable across the two samples, suggesting that the moderation finding does not simply reflect the large share of zero-price adjustments in the data. That said, the bookbuilding subsample is small ($N \approx 49$), so these stable estimates should be treated as supporting rather than conclusive, to our moderation finding.

Decomposition Window and Underwriter Fixed Effects

Hanley and Hoberg (2010) report that the standard versus informative content decomposition is robust to the choice of recent IPO peer window. We verify this for

Sweden by re-estimating the decomposition with windows of 60, 90 (baseline), 120, and 180 days and re-fitting the headline IR and ΔP regressions on each. Pairwise correlations between the resulting content variables exceed 0.93 across all window pairs, confirming that the decomposition is not sensitive to the exact window length. The headline Panel C Risk Factors IR results are similarly stable (the Stand × EU Reg coefficient remains around −0.53 to −0.55 and significant at the 1% level across all four windows), further supporting the robustness of the moderation finding to this methodological choice. Full results are in Table 28, Appendix G.

Underwriter handling is similarly robust: replacing the continuous lead-underwriter market-share control with fixed effects for the top 5 or top 10 most active underwriters shifts coefficients only marginally and leaves all significance patterns unchanged. Full results are in Table 29, Appendix G. With 269 IPOs spread across 58 underwriters, the average underwriter handles fewer than five IPOs. Most underwriters in the sample will therefore have only one or two deals, meaning per-underwriter fixed effects would absorb nearly as many degrees of freedom as there are observations, leaving insufficient variation to identify the content coefficients reliably. The continuous market-share control achieves the same goal – proxying for underwriter quality and reputation – without this collinearity problem.

Sample Composition.

The post-Regulation period coincides with a period of historically high IPO activity in Sweden, peaking in 2021 with 44 listings in the second quarter alone. To verify that the conclusions in Sections 6.1 and 6.2 are not driven by a single year, we drop each calendar year in turn and re-estimate the headline IR models. The leave-one-year-out coefficient path is shown in Figure 2.

Table 14: Robustness Check — Drop-2021 (IPO Boom Year)

Specification	Variable	Sample	Coef.	t	N
Total IR (full)	Standard	Full	+0.226	+0.84	251
		Excl. 2021	+0.295	+1.50	179
	Informative	Full	+0.221	+0.82	251
		Excl. 2021	+0.282	+1.15	179
Risk IR (full)	Standard	Full	+0.156	+1.08	249
		Excl. 2021	+0.236**	+2.01	177
	Informative	Full	−0.145	−1.21	249
		Excl. 2021	+0.002	+0.02	177

Notes: OLS coefficients (Equation 8) with t-statistics in parentheses, full controls, year and industry fixed effects, and two-way industry × year clustered standard errors. The Risk Standard coefficient turns significant when 2021 is excluded (+0.236**), consistent with the 2021 IPO boom compressing cross-sectional variation in the full sample. The full leave-one-year-out coefficient path is plotted in Figure 2. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

The leave-one-year-out path confirms that no single year is responsible for our null findings on individual coefficients. The 2021 boom year modestly influences but does not drive the results: dropping it shifts the total document standard content estimate downwards but leaves it insignificant, while in the Risk Factors section strengthens the standard content coefficient to marginal significance, suggesting 2021 dilutes rather than inflates that relationship (Table 14). One exception worth noting is the 2015 dip in Risk Factors informative content, where dropping either year produces a negative and significant coefficient. Given that these are years with

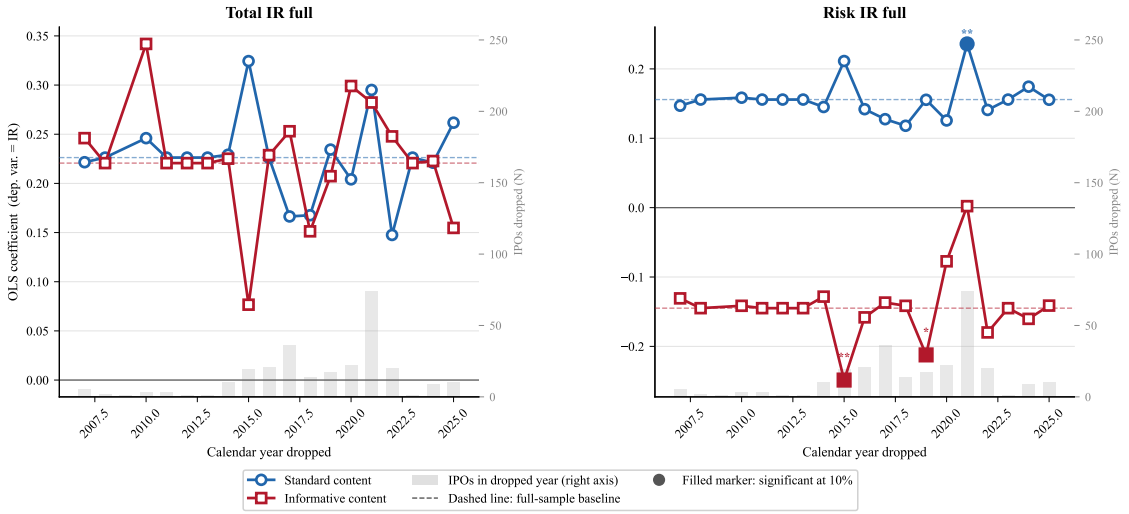


Figure 2: Stability of Content Coefficients: Leave-One-Year-Out Analysis (Total IR).

Notes: OLS regressions (Equation 8). Each point shows the coefficient on Standard and Informative content for Total IR (full controls, two-way industry × year clustered standard errors) when the indicated calendar year is dropped from the sample. Shaded bands denote 95% confidence intervals. Results for the Risk Factors section are reported in Table 14.

relatively few IPOs but high gross proceeds (Figure 4), this likely reflects sample composition rather than anything systematic. Taken together, the leave-one-year-out analysis supports the robustness of the main findings: the null on individual content-pricing coefficients is stable across all year exclusions, but our moderation finding is sensitive to exclusion of the boom year.

Summary of Robustness Checks

The robustness tests broadly supports the main findings. The null on individual content-pricing coefficients is stable across all tests – no specification produces a significant result and sign changes are rare. The Regulation moderation finding survives most tests but is sensitive to the 2021 boom, which amplifies the Total Document and Risk Factors results; the Summary section remains significant regardless. The decision to use IR over ΔP is validated by the bookbuilding subsample check. The one finding that does not survive is the EU Growth effect on informative content, which disappears under PCA dimensionality reduction and should therefore be interpreted with caution.

7. Conclusion

This study aims to answer whether the EU Prospectus Regulation (EU) 2017/1129 ("Regulation") changed the informational composition of Swedish IPO prospectuses, and whether this affects first-day pricing. The headline answer is that the Regulation did not change what issuers disclosed, but appears to have changed how the market prices a given unit of disclosure.

Prediction 1, that the Regulation would increase the degree of informative content in prospectuses, receives mixed support. For the average post-Regulation prospectus, the Regulation itself is not associated with a robust change in either stan-

dard or informative content once firm characteristics are controlled for. However, the baseline results show a positive association between the EU Growth Prospectus format and informative content, suggesting that the lighter prospectus regime may have changed disclosure composition for the smaller issuers. Though, this finding should be interpreted with caution, as it disappears under PCA dimensionality reduction and may partly reflect changes in regulatory vocabulary rather than a robust increase in firm-specific disclosure. Prediction 2, being that informative content would reduce underpricing, yields an unconditional null result, and the 8% underpricing reduction Hanley and Hoberg (2010) document does not appear in the Swedish setting. However, the joint Wald F-test shows that the content-pricing relationship differs across regulatory regimes. This suggests that the Regulation did not change the overall level of underpricing, but changed how the market values different types of disclosure. Post-Regulation, a unit of standard content appears to be priced more positively – in other words, the cost of template language has increased.

In contrast to Hanley and Hoberg (2010), who find an unconditional 8% reduction in underpricing, we cannot find a similar pattern in Sweden. Point estimates are consistent with their framework – standard content loads positively on initial returns and informative content negatively – but the confidence intervals are wide and the magnitudes are difficult to interpret precisely. This is likely due to Sweden's institutional setting. The Swedish IPO market is characterized by a smaller, more concentrated investor base, a higher prevalence of fixed-price offers accounting for 77% of zero price revisions, and a substantially smaller sample of 269 IPOs compared to the 1,750 in Hanley and Hoberg (2010). This setting provides a potential explanation for the muted content-pricing relationship. In a market like Sweden, where a small number of repeat investors hold most of the relevant private information, the bookbuilding process can extract that information effectively – as Cornelli and Goldreich (2001) document. In other words, the price is largely set through direct interaction between the issuer and a familiar group of investors, leaving less work for the prospectus to do. This is the key structural difference from the US setting: in Sweden, textual disclosure matters less at the margin because the bookbuilding process has already done much of the heavy lifting before the offer price is set. Together, these features help explain why our results look different from Hanley and Hoberg (2010) – the theory was built on assumptions that fit the US market better than Sweden, and when these assumptions cease to hold, a weaker content-pricing relationship is what we would expect to find.

One contribution of this paper is therefore that it is a replication of the textual methodology of Hanley and Hoberg (2010) outside the US, showing that the standard–informative decomposition can be applied in a non-US regulatory environment, even if the magnitudes differ. The regulatory setting further allows the decomposition to be used as a before–after test of whether harmonised disclosure rules reshape the content–pricing link. More distinctly, the paper shows that content composition and the market's pricing of that content can move independently of each other – the Regulation did not significantly change what issuers wrote, but it changed how much that writing mattered for pricing.

8. Limitations and Suggestions for Future Research

There are limitations of this study that give directions for future research. One central limitation is geographic scope. The Regulation came into force across all 27 member states in July 2019, meaning the studied regulatory shift is not unique to Sweden. Replicating the Hanley and Hoberg (2010) decomposition for other, larger markets such as Germany, France, and the Netherlands would directly test whether our findings are specific to Sweden or reflect an EU-wide regulatory effect. A cross-country study would also substantially increase statistical power beyond what a single-country sample can provide.

A second direction concerns topic-level composition of prospectus content. This paper constructs topic-specific word lists following Hanley and Hoberg's (2010) categorization of content into valuation, accounting, marketing, and due diligence language, but these are not brought into the regression analysis. Including them would make it possible to test more precisely where the moderation comes from – for instance, whether it is concentrated in risk factor language or distributed evenly across sections.

On the topic of moderation, a further limitation is that, while this paper finds that the Regulation changed how content is priced, it cannot fully explain why. The most intuitive explanation is that the Regulation made boilerplate language easier to identify, so investors began treating it as a signal that the issuer is not being forthcoming – but this remains an interpretation rather than a directly tested finding. Future research could examine this more directly by measuring how specific or generic the risk factors in a prospectus are, for instance by counting uncertainty-related words following Loughran and McDonald (2011) or tracking how often concrete company-specific details appear, and testing whether these measures predict underpricing.

Additionally, this paper argues that Sweden's concentrated investor base reduces the marginal informational value of prospectus disclosure, drawing on Cornelli and Goldreich (2001), but this hypothesis is never directly tested. With allocation data, for example from post-IPO holdings, future research could make the proposed mechanism more concrete by linking investor concentration to the degree to which prospectus content moves the first-day price.

As for AI, its rise serves as a motivation for this paper (Financial Times, 2025), but it is never tested whether the difference between human-written and AI-generated content plays a role. A future study could compare AI-drafted sections to human-drafted equivalents and ask whether AI-generated content is systematically more standard or more informative in the Hanley and Hoberg (2010) sense – and how that maps to pricing outcomes. Given that the post-Regulation environment appears to price standard language more harshly, the answer has direct implications for how such tools should be designed.

References

Abrahamson, Mark, Tim Jenkinson, and Howard Jones. 2011. "Why Don't U.S. Issuers Demand European Fees for IPOs?" *The Journal of Finance* 66 (6):2055–2082.

Aggarwal, Reena, Nagpurnanand R. Prabhala, and Manju Puri. 2002. "Institutional Allocation in Initial Public Offerings: Empirical Evidence." *The Journal of Finance* 57 (3):1421–1442.

Andersson, Jan. 2009. "Sweden." In *Prospectus for the Public Offering of Securities in Europe*, edited by Dirk Van Gerve. Cambridge: Cambridge University Press, 195–202. https://doi.org/10.1017/CBO9780511581441.014.

Benveniste, Lawrence M. and Paul A. Spindt. 1989. "How Investment Bankers Determine the Offer Price and Allocation of New Issues." *Journal of Financial Economics* 24 (2):343–361.

Carter, Richard and Steven Manaster. 1990. "Initial Public Offerings and Underwriter Reputation." *The Journal of Finance* 45 (4):1045–1067.

Chang, Yunzhi, Alexander Ljungqvist, and Charles Tseng. 2023. "Endogenous Information Acquisition and Analyst Forecasting." *Review of Financial Studies* 36 (7):3009–3055.

Cornelli, Francesca and David Goldreich. 2001. "Bookbuilding and Strategic Allocation." *The Journal of Finance* 56 (6):2337–2369.

Dalhäll, Andreas. 2025. "Starkt IPO-år: Stockholmsbörsen sticker ut globalt." EY Sweden. December 22. https://www.ey.com/sv_se/newsroom/2025/12/starkt-ipo-ar-stockholmsborsen-sticker-ut-gl.

European Commission. 2004. "Commission Regulation (EC) No 809/2004 Implementing Directive 2003/71/EC of the European Parliament and of the Council." Tech. Rep. L 149, Official Journal of the European Union.

European Parliament and Council. 2017. "Regulation (EU) 2017/1129 of the European Parliament and of the Council on the Prospectus to be Published when Securities are Offered to the Public or Admitted to Trading on a Regulated Market." Tech. rep., Official Journal of the European Union.

Financial Times. 2025. "Goldman Sachs CEO Says AI Can Compile 95% of an IPO Prospectus." https://www.ft.com/content/4f20fbb9-a10f-4a08-9a13-efa1b55dd38a.

Hanley, Kathleen Weiss and Gerard Hoberg. 2010. "The Information Content of IPO Prospectuses." *Review of Financial Studies* 23 (7):2821–2864.

Loughran, Tim and Bill McDonald. 2011. "When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks." *The Journal of Finance* 66 (1):35–65.

Megginson, William L. and Kathleen A. Weiss. 1991. "Venture Capitalist Certification in Initial Public Offerings." *The Journal of Finance* 46 (3):879–903.

Nasdaq. 2024. "Nasdaq Launches the Nasdaq Stockholm IPO Pulse: Tripled IPO Activity Since Recent Low." Press release. https://www.nasdaq.com/press-release/nasdaq-launches-nasdaq-stockholm-ipo-pulse-tripled-ipo-activity-recent-low-2024-08-27.

OECD. 2026. "The Swedish Equity Market: Assessment and Policy Recommendations 2026." Tech. rep., Organisation for Economic Co-operation and Development. https://doi.org/10.1787/0db1ebf9-en.

Rock, Kevin. 1986. "Why New Issues are Underpriced." *Journal of Financial Eco-*

nomics 15 (1–2):187–212.

SEB. 2026. "Marknaden och kurslistor – Börsen idag." https://seb.se/marknaden-och-kurslistor/borsen-idag. Accessed February 2026.

Swedish Parliament. 1991. "Lag (1991:980) om handel med finansiella instrument." Tech. rep., Svensk författningssamling. https://www.riksdagen.se/sv/dokument-och-lagar/dokument/svensk-forfattningssamling/lag-1991980-om-handel-med-finansiella_sfs-1991-980/.

———. 2005. "Lag (2005:833) om ändring i lagen (1991:980) om handel med finansiella instrument." Tech. rep., Svensk författningssamling. https://rkrattsdb.gov.se/SFSdoc/05/050833.PDF.

———. 2019. "Lag (2019:414) med kompletterande bestämmelser till EU:s prospektförordning." Tech. rep., Svensk författningssamling.

White & Case. 2024. "Global IPOs 2024: Sweden and CIS Opportunities." https://www.whitecase.com/insight-our-thinking/global-ipos-2024-sweden-cis-opportunities.

Appendices

A. Variable Descriptions

IPO size at filing Expected offer size in MSEK at the prospectus stage, calculated as the prospectus offer price, or midpoint of the indicated price range, multiplied by shares offered. Sourced from prospectus and S&P Capital IQ.

Gross proceeds Realized capital raised in the IPO, measured in MSEK. Sourced from S&P Capital IQ, supplemented manually from the prospectus when missing.

Firm age Years between founding/incorporation date and IPO transaction date. Transaction date is from S&P Capital IQ; founding dates are from S&P Capital IQ, Bolagsverket, or company websites when missing. Regressions use log firm age.

Tech dummy Equals 1 for firms classified as technology-related based on S&P Capital IQ industry group and topic fields. Firms are classified as tech when the industry group or topic fields indicate software, digital platforms, IT services, hardware, biotechnology, medtech, or other technology-intensive activities.

PE dummy Equals 1 if the issuer is private-equity backed at the time of the IPO. Sourced from S&P Capital IQ sponsor/ownership data.

Regulated market dummy Equals 1 for listings on regulated markets, including Nasdaq Stockholm and NGM Equity, and 0 for MTFs such as Nasdaq First North, Spotlight, and NGM SME. Sourced from S&P Capital IQ and prospectus venue information.

Financial sponsor dummy Equals 1 if any financial sponsor is involved, including PE, venture capital, investment companies, or other financial owners. Sourced from S&P Capital IQ.

EU Regulation dummy Equals 1 for prospectuses filed on or after 21 July 2019, the full application date of the EU Prospectus Regulation. Filing dates are from prospectus file names and FI's register.

EU growth dummy Equals 1 for prospectuses classified as EU Growth Prospectuses in FI's document type field ("Dokumenttyp": "Tillväxtprospekt"). EU Growth prospectuses are a subset of post-Regulation prospectuses.

Lead UW market share Lead underwriter's share of total Swedish IPO gross proceeds in the IPO calendar year. Calculated from S&P Capital IQ gross proceeds and manually cleaned underwriter names. Each underwriter's market share in year t is their total credited proceeds divided by the total proceeds across all underwriters in year t.

OMX30 return Cumulative OMX30 return over the 30 calendar days prior to the transaction date, normalised to a standardised 30-day window. Closing prices are from S&P Capital IQ. Defined as:

$$\text{OMX30 Return}_i = \left(\frac{\text{OMX30}(t_i)}{\text{OMX30}(t_i - 30)} \right)^{\frac{30}{30-d_i}} - 1$$

where t_i is the transaction date of IPO i; $\text{OMX30}(t_i)$ is the index level on the

transaction date; OMX30($t_i - 30$) is the index level on the nearest available trading day to 30 calendar days prior; and d_i is the number of calendar days between that nearest trading day and the exact target date $t_i - 30$. The exponent $\frac{30}{30-d_i}$ corrects for non-trading days, ensuring the return is always expressed over a standardised 30-calendar-day window.

B. Section Descriptions

Summary Provides a condensed overview of the offering and serves as the primary marketing tool used by underwriters to attract investors.

Risk Factors Lists the material risks specific to the issuer and the offering and was subject to the most prescriptive changes under the 2019 EU Prospectus Regulation.

Background and Motive Explains why the firm is going public and how it intends to use the proceeds.

Business Description Covers the issuer's operations, products, markets, and competitive position, constituting the second-largest section of the prospectus on average.

Financial Information Contains historical financial statements and related disclosures, and is the largest section on average, dominated by accounting vocabulary.

Other Content Captures all remaining mandatory disclosures not belonging to the five sections above (including governance, ownership structure, and legal matters) and serves as a residual category, ensuring that all six sections sum to one.

C. Robustness Cell Definitions

The robustness checks stress-tests six headline results, referred to as cells C1–C6, that collectively span the empirical claims of sections 6.1 and 6.2. Each cell corresponds to a specific specification, section, and dependent variable as follows:

C1 Re-estimates Total IR at step (1) (content and fixed effects only, without controls) on the Total Document, defending the baseline content–pricing relationship reported in Table 5, column (1).

C2 Re-estimates Total IR at step (3) (with the full set of firm and market controls) on the Total Document, defending the preferred pricing specification reported in Table 5, column (3).

C3 Re-estimates Risk IR at step (3) (with the full set of controls) on the Risk Factors section, defending the section-level pricing result reported in Table 5, column (3).

C4 Re-estimates the Total ΔP Panel A baseline on the Total Document, defending the price revision results reported in Table 6, Panel A.

C5 Re-estimates the Risk ΔP Panel C EU-regime interaction terms on the Risk Factors section, defending the moderation results reported in Table 6, Panel C.

C6 Re-estimates the EU Regulation and EU Growth coefficients on total standard

and informative content, defending the composition results reported in Table 4.

D. Word Lists

We use the following word lists to assess the type of topical content. The word lists are Swedish-language equivalents of those employed by Hanley and Hoberg (2010) and are constructed together with Claude Opus 4.6 (Anthropic) from publicly available Swedish institutional sources. All words are processed using a Swedish Snowball stemmer, and articles, conjunctions, personal pronouns, abbreviations, and compound-word fragments are excluded. The final vocabulary consists of 3,496 unique word roots drawn from 4,336 source phrases across ten content and tone categories. Words were further validated against the 269 Swedish IPO prospectuses filed between 2006 and 2026: phrases whose stemmed roots are absent from the IPOs were reviewed individually; English-only terms, pure abbreviations, and acronyms with no presence were removed, while Swedish-language terms that could plausibly appear in prospectuses and carry informational content were retained.

Produktmarknad (Product Market). All words appearing in Swedish SNI (*Svensk Näringsgrensindelning*) industry classification descriptions as published by Statistics Sweden (SCB), supplemented with sector-specific production, logistics, technology, and distribution terminology. Financial, accounting, and legal terms are excluded. 654 source phrases.
SNI 2007 branch descriptions: www.scb.se/dokumentation/klassifikationer-och-standarder/sni

Redovisning (Accounting). All words appearing in the Swedish Annual Accounts Act (*Årsredovisningslagen*, ÅRL), the Swedish Accounting Standards Board (*Bokföringsnämnden*, BFN) normative guidance, and IFRS terminology as applied under Swedish law, including balance sheet items, income statement line items, cash flow terminology, audit concepts, and financial ratios. 534 source phrases.
Årsredovisningslagen (1995:1554): https://www.riksdagen.se/sv/dokument-och-lagar/dokument/svensk-forfattningssamling/arsredovisningslag-19951554_sfs-1995-1554/
Bokföringsnämndens normgivning: https://www.bfn.se/om-bokforingsnamnden/var-normgivning/
IFRS as adopted by the EU: https://www.efrag.org/en/financial-reporting

Juridik (Legal). Words from Swedish corporate law, commercial law, securities law, insolvency law, employment law, competition law, data protection law, and capital markets regulation. Sources include the Swedish Companies Act (*Aktiebolagslagen*), the Financial Instruments Trading Act (*Lag om handel med finansiella instrument*), the Swedish Financial Supervisory Authority (Finansinspektionen) regulatory vocabulary, and standard Swedish legal dictionaries. 513 source phrases.
Aktiebolagslag (2005:551): https://www.riksdagen.se/sv/dokument-och-lagar/dokument/svensk-forfattningssamling/aktiebolagslag-2005551_sfs-2005-551/
Finansinspektionens regelbok: https://www.fi.se/sv/bank/regler/
Domstolsverkets juridiska ordlista: https://www.domstol.se/om-sveriges-domstolar/for-dig-som-aktor-i-domstol/stod-for-aktorer-i-domstol/ordlistor/juridisk-ordlista/

Företagsstrategi (Corporate Strategy). Merged vocabulary from Swedish-language editions of leading corporate strategy textbooks, covering competitive advantage, strategic planning, mergers and acquisitions, growth strategy, organisational design, digitalisation, and innovation strategy. 427 source phrases.
Hill, C. W. L. and Jones, G. R., Strategic Management (Swedish ed.)
Ross, S., Westerfield, R., and Jaffe, J., Corporate Finance (Swedish ed.)

Patent & Immateriella rättigheter (Patent & IP). Words from Swedish and European patent law, trademark registration, copyright, trade secret, and intellectual property licensing terminology. Sources include the Swedish Intellectual Property Office (PRV) glossary and standard IP law references. 289 source phrases.
PRV:s ordlista för immaterialrätt: https://www.prv.se/sv/kunskap-och-stod/vanliga-ord-och-begrepp/
European Patent Office glossary: https://www.epo.org/en/service-support/glossary

Marknadsföring (Marketing). Words from Swedish marketing practice covering market segmentation, branding, pricing, distribution channels, digital marketing, customer acquisition and retention, and marketing analytics. 442 source phrases.
Sveriges Marknadsförbunds terminologi: www.sverigesmarknadsforbund.se
Swedish marketing textbook glossaries (Kotler et al., Swedish ed.)

Bolagsstyrning (Corporate Governance). Words from the Swedish Corporate Governance Code (*Svensk kod för bolagsstyrning*), covering board composition, audit and remuneration committees, shareholder rights, ownership structure, internal control, risk management, and sustainability reporting. 381 source phrases.
Svensk kod för bolagsstyrning: https://bolagsstyrning.se/Userfiles/Koden/Dokument/svensk_kod_bolagsstyrning_justerad_SVE_211220.pdf
Finansinspektionens bolagsstyrningsregler: https://www.fi.se/sv/bank/regler/

Värdering (Valuation). Words from Swedish and international corporate valuation practice, including discounted cash flow, relative valuation multiples, weighted average cost of capital, enterprise and equity value concepts, and transaction analysis. Sources include FAR's financial terminology publications and standard valuation handbooks used in Swedish practice. 352 source phrases.
FAR:s ordlista för redovisning och revision: www.far.se
International Glossary of Business Valuation Terms (Swedish practice adaptation)

We also consider the following tone word lists, which are analogously defined.

Positiv ton (Positive tone). A list of Swedish positive and optimistic words relevant to IPO prospectus language, covering expressions of strength, growth, market leadership, innovation, profitability, stability, and value creation. Constructed by analogy with the Harvard General Inquirer positive word list, adapted to Swedish prospectus language conventions. 377 source phrases.
Adapted from: SenSALDO Swedish sentiment lexicon, Språkbanken Text, University of Gothenburg.
Reference: Harvard General Inquirer (English): www.wjh.harvard.edu/inquirer/homecat.htm

Negativ ton (Negative tone). A list of Swedish negative and risk-signalling words relevant to IPO prospectus language, covering risk, uncertainty, loss, competitive pressure, regulatory burden, operational vulnerability, and adverse market

conditions. Constructed by analogy with the Harvard General Inquirer negative word list, adapted to Swedish prospectus language conventions. 367 source phrases. *Adapted from: SenSALDO Swedish sentiment lexicon, Språkbanken Text, University of Gothenburg.*
Reference: Harvard General Inquirer (English): www.wjh.harvard.edu/inquirer/homecat.htm

E. Cosine Similarity

This appendix explains how we compute the pairwise document similarity between prospectuses i and j. The procedure closely follows Hanley and Hoberg (2010), adapted for Swedish-language prospectuses. The same method applies to any document section, in which case the result is referred to as section similarity.

To measure the similarity between prospectuses i and j, we take the dot product of their normalized word vectors divided by the product of their vector lengths. This quantity is known as the cosine similarity because it measures the cosine of the angle between the two vectors on a unit sphere. Formally, following equation (A1) in Hanley and Hoberg (2010):

$$\text{Document Similarity}_{\text{tot},ij} = \frac{\widetilde{\mathbf{w}}_{\text{tot},i} \cdot \widetilde{\mathbf{w}}_{\text{tot},j}}{\|\widetilde{\mathbf{w}}_{\text{tot},i}\| \, \|\widetilde{\mathbf{w}}_{\text{tot},j}\|} \tag{10}$$

Because all elements of the word vectors are non-negative, the cosine similarity is bounded in the interval $[0, 1]$. A value of 1 indicates that the two prospectuses have an identical relative word distribution, while a value of 0 indicates that they share no vocabulary at all. In practice, values closer to 1 reflect more similar prospectus content.

The cosine similarity is computed for the total document vector as well as for each of the six sections separately: Prospectus Summary, Risk Factors, Background & Motive, Business Description, Financial Information, and Other. With a sample of $N = 269$ prospectuses, this yields $N(N-1)/2 = 35{,}778$ unique prospectus pairs per word vector.

After constructing the per-prospectus word-count matrices, the normalized vectors are unit-normalized (divided by their Euclidean norm), and the full similarity matrix is obtained as the matrix product of unit-normalized row vectors, clipped to $[0, 1]$ to correct for floating-point errors. The procedure is equivalent to eq. (10) and produces a symmetric $N \times N$ similarity matrix.

One adaptation relative to Hanley and Hoberg (2010) concerns the stemmer. They use English word roots identified via Webster.com. Because our prospectuses are written in Swedish, we implement a Swedish Snowball stemmer, which applies a sequence of suffix-stripping rules calibrated to Swedish morphology.

F. Principal Component Analysis: Methodology

This appendix explains the methodology behind our PCA model and how it is structured. The methodology is based on Jolliffe (2002)[14] and was compiled using Claude Opus 4.6 (Anthropic).

[14] Jolliffe, I. T. (2002). *Principal Component Analysis* (2nd ed.). Springer, New York.

PCA reduces the dimensionality of multivariate data by transforming correlated variables into a new set of uncorrelated variables, known as principal components (PC). Each PC is a linear combination of the original variables, constructed such that the first PC captures the maximum variance in the data, the second PC captures the maximum remaining variance subject to being orthogonal to the first, and so on. In our setting, PCA is applied to the $269 \times 3{,}496$ matrix of L^1-normalized word vectors to reduce the vocabulary dimension from 3,496 to k, where k is chosen to retain a target share of total variance.

Setup. Let $\mathbf{X} \in \mathbb{R}^{N \times V}$ denote the matrix of L^1-normalized word vectors, where $N = 269$ is the number of IPO prospectuses and $V = 3{,}496$ is the vocabulary dimension. Each row $\mathbf{x}_i \in \mathbb{R}^V$ is the normalized word vector for prospectus i, satisfying $\sum_{j=1}^{V} x_{ij} = 1$ and $x_{ij} \geq 0$.

We first mean-centre the data by subtracting the column means:

$$\tilde{\mathbf{X}} = \mathbf{X} - \mathbf{1}\bar{\mathbf{x}}^\top \tag{11}$$

where $\bar{\mathbf{x}} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{x}_i \in \mathbb{R}^V$ is the column-wise mean vector and $\mathbf{1} \in \mathbb{R}^N$ is a vector of ones.

Eigendecomposition of the covariance matrix. The sample covariance matrix of the word vectors is:

$$\mathbf{S} = \frac{1}{N-1}\tilde{\mathbf{X}}^\top\tilde{\mathbf{X}} \in \mathbb{R}^{V \times V} \tag{12}$$

PCA seeks the orthonormal directions $\mathbf{w}_1, \mathbf{w}_2, \ldots \in \mathbb{R}^V$ that successively maximise projected variance. The first PC direction solves:

$$\mathbf{w}_1 = \arg\max_{\|\mathbf{w}\|=1} \mathbf{w}^\top\mathbf{S}\mathbf{w} \tag{13}$$

The solution is the eigenvector of $\mathbf{S}$ corresponding to its largest eigenvalue λ_1. More generally, the ℓ-th PC direction $\mathbf{w}_\ell$ is the eigenvector corresponding to the ℓ-th largest eigenvalue λ_ℓ, subject to $\mathbf{w}_\ell^\top\mathbf{w}_m = 0$ for all $m < \ell$.

The full eigendecomposition of the covariance matrix is:

$$\mathbf{S} = \mathbf{W}\mathbf{\Lambda}\mathbf{W}^\top \tag{14}$$

where $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_V] \in \mathbb{R}^{V \times V}$ is the matrix of eigenvectors (the *loading matrix*) and $\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_V)$ is the diagonal matrix of eigenvalues in descending order, $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_V \geq 0$.

In practice, since $N \ll V$, the covariance matrix $\mathbf{S}$ has rank at most $N-1 = 268$, meaning at most 268 eigenvalues are strictly positive. The computation is therefore performed via the Singular Value Decomposition (SVD) of $\tilde{\mathbf{X}}$ rather than directly on $\mathbf{S}$:

$$\tilde{\mathbf{X}} = \mathbf{U}\mathbf{\Sigma}\mathbf{W}^\top \tag{15}$$

where $\mathbf{U} \in \mathbb{R}^{N \times N}$ and $\mathbf{W} \in \mathbb{R}^{V \times V}$ are orthogonal matrices and $\mathbf{\Sigma} \in \mathbb{R}^{N \times V}$ is a rectangular diagonal matrix of singular values $\sigma_1 \geq \sigma_2 \geq \cdots \geq 0$. The eigenvalues of $\mathbf{S}$ relate to the singular values by $\lambda_\ell = \sigma_\ell^2/(N-1)$.

Dimensionality reduction. To reduce the vocabulary dimension from V to k, we retain only the first k eigenvectors, forming the truncated loading matrix $\mathbf{W}_k = [\mathbf{w}_1, \ldots, \mathbf{w}_k] \in \mathbb{R}^{V \times k}$. The projected (PCA-compressed) word vectors are then:

$$\mathbf{Z} = \tilde{\mathbf{X}}\mathbf{W}_k \in \mathbb{R}^{N \times k} \tag{16}$$

Each row $\mathbf{z}_i \in \mathbb{R}^k$ is the representation of prospectus i in the k-dimensional PC space. The projection is lossless along the retained directions and discards variance associated with the $V - k$ dropped eigenvectors.

The proportion of total variance explained by the first k components is:

$$\text{PVE}(k) = \frac{\sum_{\ell=1}^{k} \lambda_\ell}{\sum_{\ell=1}^{V} \lambda_\ell} = \frac{\sum_{\ell=1}^{k} \sigma_\ell^2}{\sum_{\ell=1}^{V} \sigma_\ell^2} \tag{17}$$

We select k such that $\text{PVE}(k)$ reaches a target threshold. In our robustness check, we test three thresholds: 80% ($k = 40$), 90% ($k = 80$), and 95% ($k = 125$), representing reductions in vocabulary dimension by factors of 87, 44, and 28, respectively.

Re-normalization and cosine similarity. The projected vectors $\mathbf{Z}$ have arbitrary scale and may contain negative entries, since the PC directions $\mathbf{w}_\ell$ are unconstrained. To restore compatibility with the cosine similarity framework used in the Hanley and Hoberg (2010) decomposition, each projected vector is L^2-normalized:

$$\tilde{\mathbf{z}}_i = \frac{\mathbf{z}_i}{\|\mathbf{z}_i\|_2} \tag{18}$$

so that $\|\tilde{\mathbf{z}}_i\|_2 = 1$ for all i. Under this normalization, the inner product between two prospectuses equals their cosine similarity:

$$\cos(\tilde{\mathbf{z}}_i, \tilde{\mathbf{z}}_j) = \tilde{\mathbf{z}}_i^\top \tilde{\mathbf{z}}_j \tag{19}$$

This is the direct analogue of the baseline framework, where cosine similarity in the original V-dimensional space is used to measure overlap in prospectus content.

Re-estimation of content measures. Having obtained the L^2-normalized projected vectors $\{\tilde{\mathbf{z}}_i\}_{i=1}^N$, the full Hanley and Hoberg (2010) decomposition is re-run identically to the baseline. For each prospectus i, let $\mathcal{R}(i)$ denote the set of recent peers (IPOs within 90 days prior) and $\mathcal{I}(i)$ the set of industry peers. The peer-average vectors in the compressed space are:

$$\tilde{\mathbf{z}}_{\text{rec},i} = \frac{1}{|\mathcal{R}(i)|} \sum_{j \in \mathcal{R}(i)} \tilde{\mathbf{z}}_j, \qquad \tilde{\mathbf{z}}_{\text{ind},i} = \frac{1}{|\mathcal{I}(i)|} \sum_{j \in \mathcal{I}(i)} \tilde{\mathbf{z}}_j \tag{20}$$

A note on the informative content mean. As noted in section 6.3, the informative content mean increases monotonically with k under PCA. This is a geometric consequence of the construction: the residual $\boldsymbol{\varepsilon}_i$ exists in a k-dimensional space, so its ℓ^1 norm is mechanically influenced by k. More formally, for a random vector $\boldsymbol{\varepsilon} \in \mathbb{R}^k$ with unit ℓ^2 norm, the expected ℓ^1 norm satisfies $\mathbb{E}\|\boldsymbol{\varepsilon}\|_1 \propto \sqrt{k}$,[15] meaning the informative content measure grows with the dimension of the retained space, even holding the true residual structure fixed. This does not affect the validity of the robustness check because the regressions use z-scored content variables, which are invariant to affine rescaling.

[15] This follows from the Cauchy–Schwarz inequality: $\|\boldsymbol{\varepsilon}\|_1 \leq \sqrt{k}\,\|\boldsymbol{\varepsilon}\|_2$.

G. Supplementary Figures & Tables

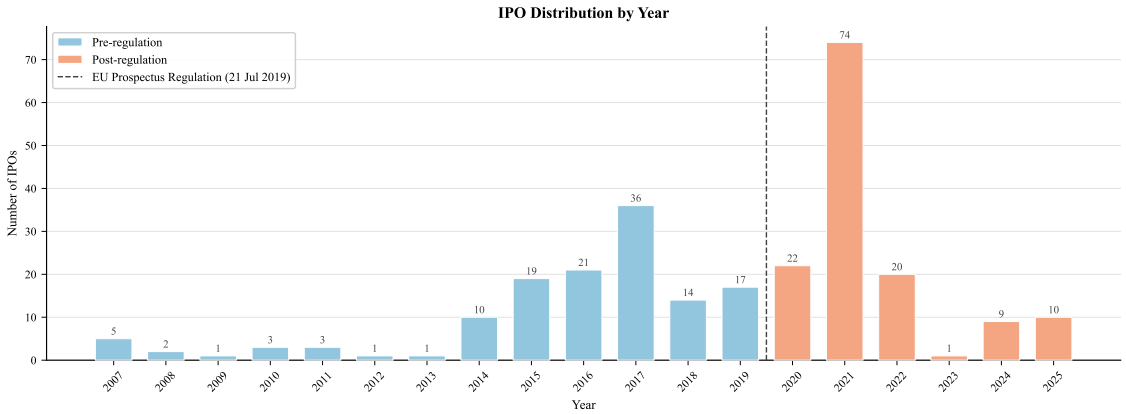


Figure 3: IPO Distribution by Year

Notes: Number of IPO prospectuses by calendar year of filing date. Blue bars indicate pre-Regulation IPOs (filed on or before 21 July 2019); red bars indicate post-Regulation IPOs (filed after 21 July 2019). The dashed vertical line marks the effective date of the EU Prospectus Regulation. Total $N = 269$.

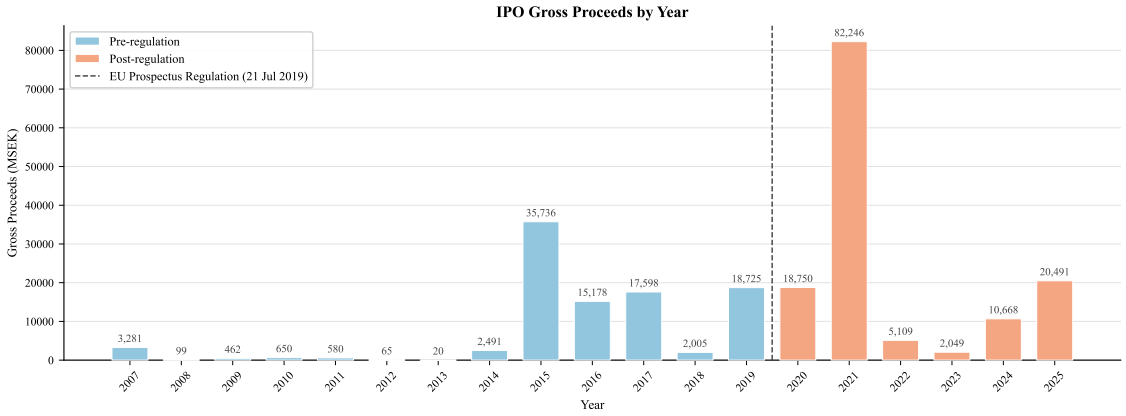


Figure 4: IPO Gross Proceeds by Year

Notes: Total gross proceeds (MSEK) by calendar year of prospectus filing date. Blue bars indicate pre-Regulation IPOs (filed on or before 21 July 2019); red bars indicate post-Regulation IPOs (filed after 21 July 2019). The dashed vertical line marks the effective date of the EU Prospectus Regulation. Total $N = 269$.

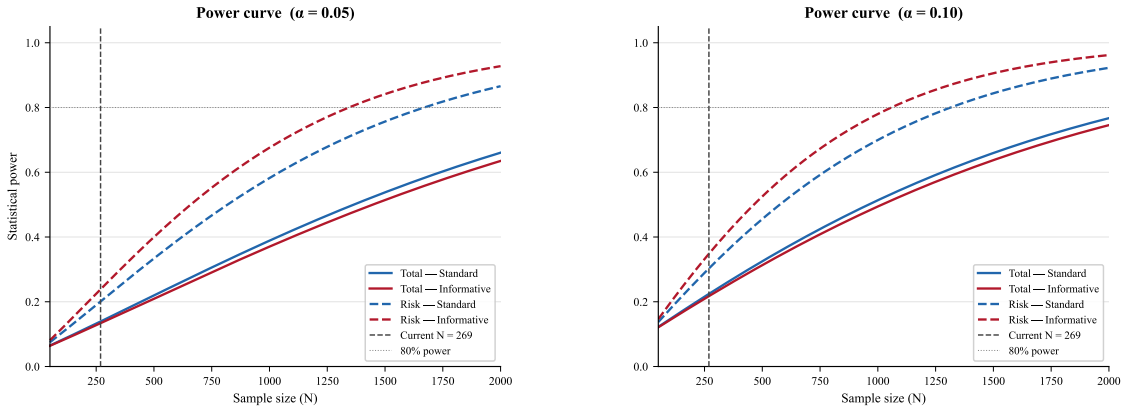


Figure 5: Statistical Power Curves

Notes: Power to detect the observed OLS coefficient at two significance levels ($\alpha = 0.05$ and $\alpha = 0.10$) as a function of hypothetical sample size. Effect sizes and residual standard deviations are taken from the full-controls IR step (3) baseline specification. Solid lines = Total document; dashed lines = Risk Factors. Blue = standard content; red = informative content. The vertical dashed line marks the actual sample size ($N = 269$) and the horizontal dotted line marks the 80% power threshold.

Table 15: Determinants of Standard Content (All Sections)

	Dep. var.: Standard content													
	Tot		*Sum*		*Risk*		*B&M*		*Bus.*		*Fin.*		*Other*	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
EU Regulation	−0.010	0.039	−0.472	−0.285	−0.079	0.029	−0.114	−0.074	−0.056	−0.114	0.003	0.071	−0.068	0.015
	(−0.16)	(1.06)	(−1.53)	(−0.68)	(−0.86)	(0.36)	(−0.75)	(−0.49)	(−0.35)	(−0.70)	(0.05)	(1.31)	(−0.54)	(0.15)
EU Growth Prospectus	0.119***	0.002	−0.005	−0.075**	0.182***	0.031	0.043	0.021	0.057	0.003	0.142**	0.037	0.050	0.019
	(2.99)	(0.05)	(−0.10)	(−2.26)	(4.54)	(0.66)	(0.97)	(0.41)	(0.93)	(0.04)	(2.13)	(0.62)	(1.46)	(0.46)
Log Firm Age	—	−0.003	—	0.066***	—	0.004	—	−0.001	—	0.014	—	0.006	—	0.009
		(−0.26)		(2.67)		(0.30)		(−0.11)		(0.74)		(0.81)		(0.99)
Log Filing Size	—	−0.043***	—	−0.038	—	−0.051***	—	0.013	—	−0.023	—	−0.043***	—	0.004
		(−5.30)		(−0.94)		(−2.95)		(0.54)		(−1.15)		(−4.44)		(0.39)
Tech Dummy	—	−0.006	—	−0.087	—	−0.009	—	−0.041	—	0.031	—	−0.028	—	0.000
		(−0.32)		(−0.69)		(−0.32)		(−1.24)		(0.96)		(−1.02)		(0.02)
PE Dummy	—	−0.034	—	0.148*	—	−0.063**	—	−0.035	—	0.048	—	−0.014	—	−0.016
		(−1.57)		(1.86)		(−2.00)		(−0.57)		(0.97)		(−0.41)		(−0.41)
Regulated Market	—	0.015	—	0.052	—	0.068	—	−0.018	—	−0.018	—	−0.011	—	−0.021
		(0.62)		(0.46)		(1.45)		(−0.38)		(−0.42)		(−0.30)		(−0.89)
Lead UW market share	—	−0.135	—	0.018	—	−0.238*	—	−0.318**	—	−0.071	—	0.005	—	−0.206**
		(−1.47)		(0.06)		(−1.69)		(−1.99)		(−0.42)		(0.06)		(−2.41)
OMX30 30-day Return	—	−0.287*	—	0.058	—	−0.244	—	0.169	—	−0.402	—	−0.520*	—	−0.258*
		(−1.72)		(0.10)		(−0.79)		(0.25)		(−1.58)		(−1.88)		(−1.90)
Log # IPOs (180d)	—	−0.004	—	0.127*	—	−0.008	—	0.001	—	−0.055	—	0.046**	—	−0.001
		(−0.20)		(1.69)		(−0.22)		(0.01)		(−1.33)		(2.21)		(−0.04)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	263	257	239	236	261	255	251	247	232	227	263	257	260	254
R^2	0.324	0.571	0.169	0.212	0.324	0.497	0.224	0.252	0.320	0.373	0.298	0.399	0.463	0.514

Notes: OLS regressions (Equation 7) with two-way industry × year clustered standard errors. Standard content = $\hat{a}_{\text{rec}} + \hat{a}_{\text{ind}}$. Odd-numbered columns include EU regime dummies and fixed effects only; even-numbered columns add the full set of controls. Pricing variables are winsorised at the 1st and 99th percentiles (baseline). t-statistics in parentheses. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table 16: Determinants of Informative Content (All Sections)

	Dep. var.: Informative content													
	Tot		*Sum*		*Risk*		*B&M*		*Bus.*		*Fin.*		*Other*	
	(15)	(16)	(17)	(18)	(19)	(20)	(21)	(22)	(23)	(24)	(25)	(26)	(27)	(28)
EU Regulation	0.041	0.049	−0.077	0.029	0.090	0.130	0.060	0.089	−0.070	−0.151	−0.019	−0.035	−0.011	−0.028
	(1.46)	(1.32)	(−0.27)	(0.08)	(1.03)	(1.25)	(0.98)	(0.99)	(−0.54)	(−1.22)	(−0.20)	(−0.59)	(−0.14)	(−0.29)
EU Growth Prospectus	0.123***	0.095***	0.129	0.063	0.136***	0.070**	0.106*	0.121	0.084	0.069	0.437***	0.344***	0.265***	0.212***
	(7.29)	(5.61)	(1.21)	(0.55)	(4.99)	(2.01)	(1.88)	(1.40)	(1.00)	(0.84)	(7.28)	(4.75)	(6.97)	(4.29)
Log Firm Age	—	−0.011**	—	0.051**	—	−0.004	—	−0.005	—	−0.024	—	−0.012	—	−0.011
		(−2.06)		(2.48)		(−0.44)		(−0.29)		(−1.18)		(−1.08)		(−0.96)
Log Filing Size	—	−0.009	—	−0.022	—	−0.027***	—	0.031	—	0.006	—	−0.041**	—	−0.019**
		(−1.27)		(−0.71)		(−3.54)		(1.09)		(0.24)		(−2.35)		(−2.18)
Tech Dummy	—	−0.001	—	−0.051	—	−0.016	—	−0.046	—	−0.027	—	0.001	—	−0.022
		(−0.09)		(−0.57)		(−0.69)		(−1.42)		(−0.68)		(0.03)		(−0.86)
PE Dummy	—	−0.033*	—	0.068	—	−0.032	—	−0.048*	—	−0.076*	—	0.012	—	0.038
		(−1.79)		(0.72)		(−1.35)		(−1.75)		(−1.68)		(0.29)		(0.56)
Regulated Market	—	−0.007	—	−0.004	—	0.020	—	0.046	—	−0.026	—	0.023	—	−0.101**
		(−0.71)		(−0.05)		(0.75)		(1.20)		(−0.78)		(0.49)		(−2.01)
Lead UW market share	—	0.008	—	−0.085	—	−0.007	—	−0.373***	—	−0.056	—	−0.084	—	−0.034
		(0.13)		(−0.34)		(−0.06)		(−2.87)		(−0.28)		(−0.68)		(−0.58)
OMX30 30-day Return	—	−0.054	—	0.351	—	−0.142	—	0.594*	—	0.233	—	−0.192	—	0.167
		(−0.39)		(0.84)		(−0.64)		(1.75)		(0.67)		(−0.45)		(0.67)
Log # IPOs (180d)	—	0.007	—	−0.005	—	0.014	—	0.026	—	−0.101*	—	0.000	—	−0.006
		(0.41)		(−0.07)		(0.52)		(0.53)		(−1.75)		(0.01)		(−0.23)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	263	257	239	236	261	255	251	247	232	227	263	257	260	254
R^2	0.535	0.575	0.147	0.181	0.370	0.411	0.111	0.161	0.138	0.191	0.515	0.581	0.415	0.469

Notes: OLS regressions (Equation 7) with two-way industry × year clustered standard errors. Informative content = $\sum_w |\varepsilon_{i,w}|$. Odd-numbered columns include EU regime dummies and fixed effects only; even-numbered columns add the full set of controls. Pricing variables are winsorised at the 1st and 99th percentiles (baseline). t-statistics in parentheses. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table 17: Content and Initial Returns (IR) — Step-by-Step Specifications (All Sections)

	Dep. var.: IR (first-day return)																				
	Tot			*Sum*			*Risk*			*B&M*			*Bus.*			*Fin.*			*Other*		
	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
Standard content	0.117	0.125	0.226	0.010	0.005	−0.002	0.134	0.131	0.156	−0.111	−0.118	−0.123	0.027	0.024	0.065	0.081	0.064	0.173	0.006	−0.004	0.003
	(0.77)	(0.69)	(0.84)	(0.14)	(0.07)	(−0.02)	(1.61)	(1.48)	(1.08)	(−1.51)	(−1.63)	(−1.63)	(0.40)	(0.31)	(0.76)	(0.48)	(0.35)	(1.07)	(0.05)	(−0.03)	(0.02)
Informative content	0.187	0.280	0.221	0.003	0.008	−0.002	−0.123	−0.104	−0.145	0.059	0.071	0.069	0.022	0.019	−0.036	0.055	0.113	0.067	0.050	0.065	0.075
	(0.79)	(0.88)	(0.82)	(0.04)	(0.11)	(−0.03)	(−1.03)	(−0.79)	(−1.21)	(0.81)	(0.90)	(0.76)	(0.19)	(0.16)	(−0.28)	(0.42)	(0.70)	(0.44)	(0.33)	(0.51)	(0.64)
EU Regulation	—	−0.136	−0.132	—	−0.082	−0.035	—	−0.106	−0.098	—	−0.127	−0.087	—	−0.238*	−0.257	—	−0.123	−0.120	—	−0.125	−0.110
		(−0.95)	(−0.84)		(−0.49)	(−0.19)		(−0.70)	(−0.57)		(−0.81)	(−0.48)		(−1.82)	(−1.62)		(−0.84)	(−0.72)		(−0.85)	(−0.67)
EU Growth (incr.)	—	−0.050	−0.030	—	−0.012	−0.025	—	−0.011	−0.002	—	−0.023	−0.037	—	−0.007	0.026	—	−0.059	−0.037	—	−0.018	−0.024
		(−0.38)	(−0.28)		(−0.12)	(−0.23)		(−0.10)	(−0.02)		(−0.27)	(−0.37)		(−0.07)	(0.27)		(−0.52)	(−0.34)		(−0.23)	(−0.27)
Log firm age	—	—	0.022*	—	—	0.018	—	—	0.018	—	—	0.015	—	—	0.013	—	—	0.018	—	—	0.020
			(1.74)			(1.39)			(1.26)			(1.25)			(1.08)			(1.38)			(1.45)
Log filing size	—	—	−0.002	—	—	−0.014	—	—	−0.009	—	—	−0.018	—	—	0.000	—	—	−0.004	—	—	−0.013
			(−0.07)			(−0.71)			(−0.44)			(−0.93)			(0.02)			(−0.18)			(−0.81)
Tech dummy	—	—	−0.076	—	—	−0.098	—	—	−0.079	—	—	−0.077	—	—	−0.111	—	—	−0.074	—	—	−0.077
			(−0.88)			(−1.10)			(−0.90)			(−0.82)			(−1.17)			(−0.86)			(−0.87)
PE dummy	—	—	0.032	—	—	0.005	—	—	0.020	—	—	0.006	—	—	0.027	—	—	0.019	—	—	0.014
			(1.21)			(0.15)			(0.65)			(0.21)			(0.95)			(0.57)			(0.49)
Regulated market	—	—	0.022	—	—	0.027	—	—	0.021	—	—	0.052	—	—	0.055	—	—	0.031	—	—	0.032
			(0.35)			(0.38)			(0.35)			(0.77)			(0.89)			(0.53)			(0.52)
Pre-file OMX30 return	—	—	0.256	—	—	0.041	—	—	0.198	—	—	0.074	—	—	0.198	—	—	0.289	—	—	0.105
			(0.89)			(0.16)			(0.78)			(0.36)			(0.66)			(0.90)			(0.30)
Lead UW market share	—	—	0.124	—	—	0.119	—	—	0.137	—	—	0.133	—	—	0.087	—	—	0.104	—	—	0.109
			(0.88)			(0.69)			(0.92)			(0.94)			(0.56)			(0.74)			(0.75)
Log # IPOs (180d)	—	—	−0.060	—	—	−0.075	—	—	−0.055	—	—	−0.060	—	—	−0.085	—	—	−0.065	—	—	−0.061
			(−0.85)			(−0.96)			(−0.74)			(−0.80)			(−0.95)			(−0.88)			(−0.84)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	257	257	251	233	233	230	255	255	249	246	246	242	228	228	223	257	257	251	254	254	248
R^2	0.139	0.144	0.186	0.156	0.157	0.187	0.130	0.132	0.175	0.144	0.147	0.173	0.104	0.112	0.163	0.137	0.142	0.188	0.131	0.134	0.176

Notes: OLS regressions (Equation 8) with two-way industry × year clustered standard errors. Dependent variable is IR (first-day return). Column (1): content and fixed effects only. Column (2): adds EU Regulation and EU Growth Prospectus dummies. Column (3): adds full controls (log firm age, log filing size, tech dummy, PE dummy, regulated market dummy, lead underwriter market share, OMX30 30-day return, log # IPOs (180d)). EU Regulation = 1 for IPOs filed on or after 21 July 2019; EU Growth (incr.) denotes the incremental effect of the EU Growth sub-regime. No content coefficient is statistically significant at conventional levels across any specification or section. Pricing variables are winsorised at the 1st and 99th percentiles (baseline). t-statistics in parentheses. Significance: $^{***}\,p<0.01$, $^{**}\,p<0.05$, $^{*}\,p<0.10$.

Table 18: Content and IPO Pricing — Panel A (Baseline, content + controls) — All Sections

	Tot			*Sum*			*Risk*			*B&M*			*Bus.*			*Fin.*			*Other*		
	ΔP	$\lvert\Delta P\rvert$	IR	ΔP	$\lvert\Delta P\rvert$	IR	ΔP	$\lvert\Delta P\rvert$	IR	ΔP	$\lvert\Delta P\rvert$	IR	ΔP	$\lvert\Delta P\rvert$	IR	ΔP	$\lvert\Delta P\rvert$	IR	ΔP	$\lvert\Delta P\rvert$	IR
Standard content	−0.231	−0.137	0.233	−0.055	−0.079	0.001	−0.088	−0.038	0.157	−0.174	−0.186	−0.117	0.024	0.056	0.071	−0.167	−0.096	0.184	−0.162	−0.168	0.005
	(−0.78)	(−0.54)	(0.82)	(−0.79)	(−1.14)	(0.02)	(−0.55)	(−0.23)	(1.11)	(−1.22)	(−1.29)	(−1.46)	(0.50)	(1.02)	(0.88)	(−0.82)	(−0.55)	(1.26)	(−0.88)	(−0.80)	(0.04)
Informative content	0.092	0.345	0.163	0.004	0.015	−0.005	−0.191***	−0.133**	−0.158	0.106	0.125	0.055	−0.005	0.009	−0.024	−0.039	0.215	0.037	0.270	0.276	0.062
	(0.39)	(1.15)	(0.63)	(0.04)	(0.15)	(−0.08)	(−2.99)	(−2.31)	(−1.38)	(0.51)	(0.58)	(0.72)	(−0.17)	(0.40)	(−0.19)	(−0.17)	(0.95)	(0.26)	(1.13)	(1.16)	(0.47)
Log firm age	−0.014	−0.015	0.021	−0.014	−0.018	0.017	−0.015	−0.019	0.018	−0.014	−0.018	0.015	−0.017	−0.014	0.014	−0.014	−0.016	0.017	−0.022	−0.026	0.020
	(−0.56)	(−0.59)	(1.58)	(−0.50)	(−0.61)	(1.32)	(−0.60)	(−0.74)	(1.25)	(−0.59)	(−0.74)	(1.24)	(−1.56)	(−1.24)	(1.10)	(−0.59)	(−0.69)	(1.30)	(−0.98)	(−1.12)	(1.42)
Log filing size	−0.060**	−0.037	0.001	−0.062*	−0.047	−0.010	−0.065***	−0.046	−0.009	−0.056**	−0.043	−0.013	−0.018**	−0.009	−0.004	−0.064**	−0.024	−0.001	−0.015	−0.002	−0.011
	(−2.33)	(−1.29)	(0.03)	(−1.93)	(−1.32)	(−0.35)	(−2.62)	(−1.57)	(−0.35)	(−2.03)	(−1.32)	(−0.52)	(−2.47)	(−0.76)	(−0.15)	(−2.04)	(−0.66)	(−0.03)	(−0.72)	(−0.08)	(−0.52)
Tech dummy	−0.053**	−0.085***	−0.076	−0.062**	−0.097***	−0.099	−0.055**	−0.086***	−0.078	−0.051**	−0.085***	−0.079	−0.032*	−0.055***	−0.103	−0.056**	−0.090***	−0.074	−0.035	−0.067**	−0.077
	(−2.04)	(−4.32)	(−0.90)	(−2.13)	(−4.11)	(−1.14)	(−2.03)	(−4.04)	(−0.92)	(−2.09)	(−4.59)	(−0.88)	(−1.77)	(−2.84)	(−1.10)	(−2.03)	(−4.60)	(−0.87)	(−1.24)	(−2.17)	(−0.91)
PE dummy	0.024	−0.009	0.032	0.039	−0.001	0.007	0.013	−0.026	0.020	0.029	−0.015	0.008	0.029	−0.003	0.025	0.024	−0.016	0.022	0.035	−0.012	0.016
	(0.35)	(−0.12)	(1.32)	(0.57)	(−0.02)	(0.20)	(0.18)	(−0.33)	(0.70)	(0.49)	(−0.25)	(0.31)	(0.74)	(−0.06)	(0.86)	(0.38)	(−0.22)	(0.77)	(0.42)	(−0.14)	(0.64)
Regulated market	−0.074	−0.105	0.028	−0.077	−0.108	0.029	−0.069	−0.105	0.026	−0.101	−0.136	0.056	−0.030	−0.059	0.068	−0.079	−0.118	0.038	−0.031	−0.064	0.037
	(−0.57)	(−0.91)	(0.46)	(−0.59)	(−0.95)	(0.42)	(−0.55)	(−0.92)	(0.45)	(−0.75)	(−1.14)	(0.89)	(−0.95)	(−1.59)	(1.11)	(−0.61)	(−1.03)	(0.66)	(−0.36)	(−0.83)	(0.60)
Pre-file OMX30 return	−1.630*	−1.539*	0.250	−1.466*	−1.417*	0.041	−1.623*	−1.553*	0.190	−1.605*	−1.554*	0.085	0.065	0.026	0.141	−1.669*	−1.517	0.286	−0.302	−0.280	0.100
	(−1.78)	(−1.73)	(0.83)	(−1.71)	(−1.69)	(0.15)	(−1.85)	(−1.80)	(0.69)	(−1.84)	(−1.82)	(0.36)	(0.16)	(0.06)	(0.44)	(−1.86)	(−1.62)	(0.87)	(−1.29)	(−1.19)	(0.26)
Lead UW market share	0.260	0.210	0.104	0.368	0.282	0.103	0.294	0.245	0.131	0.295*	0.243	0.106	0.090	0.061	0.085	0.305	0.230	0.079	0.071	0.024	0.091
	(1.28)	(0.87)	(0.61)	(1.58)	(1.05)	(0.48)	(1.33)	(0.88)	(0.74)	(1.77)	(1.08)	(0.60)	(1.21)	(0.63)	(0.51)	(1.59)	(1.02)	(0.44)	(0.55)	(0.13)	(0.54)
Log # IPOs (180d)	−0.101	−0.124*	−0.054	−0.088	−0.105	−0.075	−0.097	−0.121*	−0.050	−0.099	−0.124	−0.056	−0.006	−0.029	−0.072	−0.092	−0.118*	−0.061	−0.041	−0.063*	−0.056
	(−1.32)	(−1.77)	(−0.78)	(−1.14)	(−1.56)	(−1.04)	(−1.34)	(−1.84)	(−0.71)	(−1.22)	(−1.65)	(−0.79)	(−0.20)	(−0.89)	(−0.87)	(−1.22)	(−1.71)	(−0.87)	(−1.25)	(−1.92)	(−0.81)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	257	257	251	236	236	230	255	255	249	247	247	242	227	227	223	257	257	251	254	254	248
R^2	0.218	0.222	0.183	0.232	0.234	0.187	0.224	0.220	0.173	0.232	0.238	0.171	0.161	0.190	0.154	0.222	0.228	0.185	0.223	0.233	0.174

Notes: OLS regressions (Equation 9, Panel A) with two-way industry × year clustered standard errors. ΔP = prospectus-to-offer price change; $|\Delta P|$ = absolute thereof; IR = first-day return. Controls: log firm age, log filing size, tech dummy, PE dummy, regulated market dummy, lead underwriter market share, OMX30 30-day return, log # IPOs (180d). Pricing variables are winsorised at the 1st and 99th percentiles (baseline). t-statistics in parentheses. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table 19: Content and IPO Pricing — Panel B (Additive EU regime dummies) — All Sections

	Tot			*Sum*			*Risk*			*B&M*			*Bus.*			*Fin.*			*Other*		
	ΔP	$\lvert\Delta P\rvert$	IR	ΔP	$\lvert\Delta P\rvert$	IR	ΔP	$\lvert\Delta P\rvert$	IR	ΔP	$\lvert\Delta P\rvert$	IR	ΔP	$\lvert\Delta P\rvert$	IR	ΔP	$\lvert\Delta P\rvert$	IR	ΔP	$\lvert\Delta P\rvert$	IR
Standard content	−0.236	−0.115	0.226	−0.061	−0.077	−0.002	−0.089	−0.040	0.156	−0.181	−0.185	−0.123	0.022	0.057	0.065	−0.158	−0.078	0.173	−0.176	−0.165	0.003
	(−0.78)	(−0.44)	(0.84)	(−0.87)	(−1.09)	(−0.02)	(−0.56)	(−0.24)	(1.08)	(−1.27)	(−1.33)	(−1.63)	(0.44)	(1.05)	(0.76)	(−0.70)	(−0.39)	(1.07)	(−0.92)	(−0.77)	(0.02)
Informative content	0.138	0.274	0.221	0.011	0.011	−0.002	−0.184***	−0.154**	−0.145	0.118	0.112	0.069	−0.002	0.004	−0.036	−0.051	0.186	0.067	0.322	0.272	0.075
	(0.54)	(0.89)	(0.82)	(0.11)	(0.11)	(−0.03)	(−2.93)	(−2.50)	(−1.21)	(0.55)	(0.51)	(0.76)	(−0.06)	(0.16)	(−0.28)	(−0.19)	(0.67)	(0.44)	(1.17)	(1.05)	(0.64)
EU Regulation	−0.104	−0.136	−0.132	−0.090	−0.117	−0.035	−0.079	−0.106	−0.098	−0.132	−0.163	−0.087	−0.025	−0.021	−0.257	−0.097	−0.115	−0.120	−0.087	−0.110	−0.110
	(−0.84)	(−1.17)	(−0.84)	(−0.55)	(−0.74)	(−0.19)	(−0.62)	(−0.86)	(−0.57)	(−1.07)	(−1.38)	(−0.48)	(−0.56)	(−0.52)	(−1.62)	(−0.69)	(−0.88)	(−0.72)	(−0.87)	(−1.18)	(−0.67)
EU Growth (incr.)	−0.024	0.067	−0.030	−0.043	0.068	−0.025	0.005	0.105	−0.002	−0.024	0.089	−0.037	−0.029	0.040	0.026	0.012	0.032	−0.037	−0.105	0.010	−0.024
	(−0.32)	(0.93)	(−0.28)	(−0.77)	(1.10)	(−0.23)	(0.08)	(1.45)	(−0.02)	(−0.38)	(1.30)	(−0.37)	(−0.97)	(0.83)	(0.27)	(0.16)	(0.35)	(−0.34)	(−1.12)	(0.12)	(−0.27)
Log firm age	−0.014	−0.017	0.022*	−0.014	−0.019	0.018	−0.015	−0.020	0.018	−0.014	−0.019	0.015	−0.016	−0.015	0.013	−0.014	−0.017	0.018	−0.021	−0.026	0.020
	(−0.52)	(−0.62)	(1.74)	(−0.47)	(−0.63)	(1.39)	(−0.60)	(−0.79)	(1.26)	(−0.56)	(−0.78)	(1.25)	(−1.54)	(−1.39)	(1.08)	(−0.61)	(−0.74)	(1.38)	(−0.89)	(−1.10)	(1.45)
Log filing size	−0.062***	−0.028	−0.002	−0.068**	−0.037	−0.014	−0.064***	−0.031	−0.009	−0.059***	−0.028	−0.018	−0.023**	−0.003	0.000	−0.062**	−0.021	−0.004	−0.028*	−0.000	−0.013
	(−3.07)	(−1.13)	(−0.07)	(−2.51)	(−1.15)	(−0.71)	(−3.30)	(−1.26)	(−0.44)	(−2.84)	(−1.12)	(−0.93)	(−2.59)	(−0.20)	(0.02)	(−2.52)	(−0.70)	(−0.18)	(−1.68)	(−0.01)	(−0.81)
Tech dummy	−0.054**	−0.089***	−0.076	−0.062**	−0.100***	−0.098	−0.056**	−0.092***	−0.079	−0.051*	−0.092***	−0.077	−0.031	−0.057***	−0.111	−0.057**	−0.091***	−0.074	−0.031	−0.069**	−0.077
	(−2.01)	(−4.19)	(−0.88)	(−2.08)	(−4.08)	(−1.10)	(−1.99)	(−4.04)	(−0.90)	(−1.96)	(−4.54)	(−0.82)	(−1.53)	(−2.74)	(−1.17)	(−2.06)	(−4.46)	(−0.86)	(−0.99)	(−2.11)	(−0.87)
PE dummy	0.023	−0.005	0.032	0.036	0.004	0.005	0.014	−0.017	0.020	0.027	−0.009	0.006	0.027	−0.000	0.027	0.025	−0.014	0.019	0.024	−0.011	0.014
	(0.34)	(−0.07)	(1.21)	(0.53)	(0.05)	(0.15)	(0.19)	(−0.21)	(0.65)	(0.46)	(−0.13)	(0.21)	(0.74)	(−0.00)	(0.95)	(0.36)	(−0.17)	(0.57)	(0.30)	(−0.12)	(0.49)
Regulated market	−0.078	−0.114	0.022	−0.080	−0.114	0.027	−0.073	−0.112	0.021	−0.107	−0.145	0.052	−0.032	−0.059	0.055	−0.083	−0.122	0.031	−0.028	−0.069	0.032
	(−0.60)	(−0.99)	(0.35)	(−0.61)	(−1.00)	(0.38)	(−0.57)	(−0.99)	(0.35)	(−0.78)	(−1.21)	(0.77)	(−0.94)	(−1.61)	(0.89)	(−0.63)	(−1.06)	(0.53)	(−0.32)	(−0.89)	(0.52)
Pre-file OMX30 return	−1.625*	−1.513*	0.256	−1.463*	−1.391*	0.041	−1.615*	−1.526*	0.198	−1.608*	−1.510*	0.074	0.062	0.040	0.198	−1.657*	−1.499	0.289	−0.315	−0.265	0.105
	(−1.75)	(−1.72)	(0.89)	(−1.67)	(−1.68)	(0.16)	(−1.84)	(−1.82)	(0.78)	(−1.82)	(−1.82)	(0.36)	(0.15)	(0.09)	(0.66)	(−1.85)	(−1.63)	(0.90)	(−1.37)	(−1.13)	(0.30)
Lead UW market share	0.276	0.189	0.124	0.397*	0.258	0.119	0.295	0.194	0.137	0.315**	0.198	0.133	0.106	0.043	0.087	0.305*	0.222	0.104	0.126	0.026	0.109
	(1.47)	(0.84)	(0.88)	(1.81)	(1.02)	(0.69)	(1.44)	(0.76)	(0.92)	(2.23)	(1.07)	(0.94)	(1.54)	(0.49)	(0.56)	(1.96)	(1.16)	(0.74)	(1.04)	(0.15)	(0.75)
Log # IPOs (180d)	−0.106	−0.130*	−0.060	−0.090	−0.110	−0.075	−0.101	−0.126*	−0.055	−0.105	−0.132*	−0.060	−0.007	−0.030	−0.085	−0.097	−0.124*	−0.065	−0.043	−0.068*	−0.061
	(−1.35)	(−1.81)	(−0.85)	(−1.16)	(−1.59)	(−0.96)	(−1.35)	(−1.85)	(−0.74)	(−1.27)	(−1.71)	(−0.80)	(−0.23)	(−0.93)	(−0.95)	(−1.22)	(−1.70)	(−0.88)	(−1.20)	(−1.95)	(−0.84)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	257	257	251	236	236	230	255	255	249	247	247	242	227	227	223	257	257	251	254	254	248
R^2	0.220	0.227	0.186	0.234	0.238	0.187	0.224	0.228	0.175	0.234	0.246	0.173	0.164	0.194	0.163	0.223	0.230	0.188	0.234	0.235	0.176

Notes: OLS regressions (Equation 9, Panel B) with two-way industry × year clustered standard errors. ΔP = prospectus-to-offer price change; $|\Delta P|$ = absolute price change; IR = first-day return. EU Regulation = 1 for IPOs filed on or after 21 July 2019; EU Growth (incr.) denotes the incremental effect of the EU Growth sub-regime. Controls: log firm age, log filing size, tech dummy, PE dummy, regulated market dummy, lead underwriter market share, OMX30 30-day return, log # IPOs (180d). Pricing variables are winsorised at the 1st and 99th percentiles (baseline). t-statistics in parentheses. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table 20: Content and IPO Pricing — Panel C (Content × EU Regulation interaction) — All Sections

	Tot			*Sum*			*Risk*			*B&M*			*Bus.*			*Fin.*			*Other*		
	ΔP	$\lvert\Delta P\rvert$	IR	ΔP	$\lvert\Delta P\rvert$	IR	ΔP	$\lvert\Delta P\rvert$	IR	ΔP	$\lvert\Delta P\rvert$	IR	ΔP	$\lvert\Delta P\rvert$	IR	ΔP	$\lvert\Delta P\rvert$	IR	ΔP	$\lvert\Delta P\rvert$	IR
Standard content	−0.044	0.061	0.301*	−0.143**	−0.124**	0.015	−0.029	−0.012	0.251***	−0.051	−0.050	−0.193**	0.005	0.010	−0.115	0.130	0.153	0.015	0.018	0.052	−0.124
	(−0.13)	(0.19)	(1.67)	(−2.59)	(−2.32)	(0.17)	(−0.24)	(−0.11)	(3.11)	(−0.63)	(−0.62)	(−1.97)	(0.06)	(0.13)	(−0.72)	(0.75)	(0.95)	(0.05)	(0.08)	(0.27)	(−0.62)
Informative content	−0.236	−0.285	0.221	0.138**	0.099*	−0.002	−0.196	−0.239**	−0.021	−0.197	−0.189	0.062	0.084	0.125*	−0.037	−0.409**	−0.395**	0.394*	0.592	0.547	0.045
	(−0.44)	(−0.55)	(0.74)	(2.36)	(1.79)	(−0.02)	(−1.56)	(−2.15)	(−0.15)	(−1.15)	(−0.98)	(0.40)	(0.88)	(1.94)	(−0.21)	(−2.30)	(−2.36)	(1.97)	(1.44)	(1.36)	(0.42)
Stand × EU Reg.	−0.226	−0.286	−0.469*	0.147*	0.126	−0.111	−0.204	−0.188	−0.537***	−0.084	−0.130	−0.040	0.007	0.014	0.003	−0.281	−0.400	0.122	−0.002	−0.093	0.430
	(−0.47)	(−0.69)	(−1.79)	(1.79)	(1.15)	(−0.96)	(−0.78)	(−0.82)	(−2.74)	(−0.61)	(−0.89)	(−0.26)	(0.09)	(0.16)	(0.01)	(−1.02)	(−1.56)	(0.34)	(−0.01)	(−0.33)	(0.91)
Inf × EU Reg.	0.365	0.630	−0.367	−0.132	−0.110	−0.119	0.071	0.141	−0.143	0.440*	0.479*	0.272*	−0.147	−0.177**	0.210	0.244	0.489	−0.804**	−0.689	−0.459	0.118
	(0.56)	(1.03)	(−0.76)	(−1.31)	(−1.13)	(−0.79)	(0.28)	(0.77)	(−0.82)	(1.81)	(1.97)	(1.65)	(−1.19)	(−2.02)	(1.22)	(0.75)	(1.57)	(−2.57)	(−1.56)	(−1.24)	(0.57)
Stand × EU Growth	−1.302	−1.380	0.669	0.079	−0.101	0.140	0.246	0.200	1.100***	−0.371	−0.297	0.375**	0.113	0.228**	0.643**	−0.304	−0.077	0.259	−0.619	−0.448	−0.374
	(−1.22)	(−1.38)	(0.67)	(0.29)	(−0.40)	(0.53)	(1.29)	(1.29)	(2.99)	(−1.04)	(−0.75)	(2.51)	(1.02)	(2.27)	(2.43)	(−0.67)	(−0.20)	(0.71)	(−1.19)	(−0.90)	(−0.95)
Inf × EU Growth	3.522*	4.339**	1.051	−0.146	−0.041	0.212	−0.150	0.250	−0.711**	0.055	−0.018	−0.334	0.022	−0.034	−0.471	0.415	0.489	0.382	0.385	0.064	−0.031
	(1.73)	(2.10)	(0.90)	(−0.57)	(−0.18)	(1.01)	(−0.27)	(0.45)	(−1.99)	(0.14)	(−0.05)	(−1.61)	(0.24)	(−0.34)	(−1.58)	(0.67)	(0.89)	(1.38)	(1.48)	(0.28)	(−0.14)
EU Regulation	−0.059	−0.176	0.557**	−0.033	−0.070	0.212	0.063	−0.025	0.521**	−0.550*	−0.587*	−0.385*	0.110	0.135*	−0.484	−0.019	−0.098	0.347	0.371	0.285	−0.636
	(−0.12)	(−0.36)	(2.12)	(−0.14)	(−0.31)	(0.84)	(0.22)	(−0.09)	(2.33)	(−1.72)	(−1.79)	(−1.82)	(1.15)	(1.94)	(−1.56)	(−0.06)	(−0.34)	(1.09)	(1.04)	(1.06)	(−1.33)
EU Growth (incr.)	−0.843**	−1.197***	−1.358**	0.067	0.215	−0.417**	−0.121	−0.322	−0.535	0.268	0.391	−0.027	−0.164	−0.154	−0.139	−0.080	−0.400	−0.526	0.327	0.492	0.354
	(−2.31)	(−3.17)	(−2.13)	(0.47)	(1.15)	(−2.14)	(−0.26)	(−0.71)	(−1.40)	(0.73)	(1.02)	(−0.14)	(−1.07)	(−0.94)	(−0.25)	(−0.21)	(−1.48)	(−1.57)	(0.72)	(1.05)	(1.02)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	257	257	251	236	236	230	255	255	249	247	247	242	227	227	223	257	257	251	254	254	248
R^2	0.268	0.306	0.209	0.245	0.244	0.208	0.228	0.235	0.227	0.257	0.266	0.184	0.175	0.217	0.199	0.243	0.266	0.215	0.274	0.270	0.181

Notes: OLS regressions (Equation 9, Panel C) with two-way industry × year clustered standard errors. ΔP = prospectus-to-offer price change; $|\Delta P|$ = absolute price change; IR = first-day return. EU Regulation = 1 for IPOs filed on or after 21 July 2019; EU Growth (incr.) denotes the incremental effect of the EU Growth sub-regime. Interaction terms test whether the content–pricing slope differs under the EU regime. Controls: log firm age, log filing size, tech dummy, PE dummy, regulated market dummy, lead underwriter market share, OMX30 30-day return, log # IPOs (180d). Pricing variables are winsorised at the 1st and 99th percentiles (baseline). t-statistics in parentheses. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table 21: Joint Wald F-Test: EU Regime Moderation of Content–Pricing Channel (All Sections)

	ΔP		$\lvert\Delta P\rvert$		*IR*	
Section	F	p	F	p	F	p
Tot	1.613	0.1725	3.679	0.0065***	5.319	0.0004***
Sum	1.061	0.3772	0.738	0.5671	3.535	0.0084***
Risk	0.661	0.6195	0.960	0.4306	3.621	0.0072***
B&M	1.721	0.1468	1.508	0.2014	2.766	0.0288**
Bus.	0.600	0.6629	3.574	0.0079***	19.446	0.0000***
Fin.	1.156	0.3317	4.903	0.0009***	6.422	0.0001***
Other	0.890	0.4710	1.465	0.2142	0.759	0.5529

Notes: Joint Wald F-test of H_0: $\beta_{\text{Stand}\times\text{EUreg}} = \beta_{\text{Inf}\times\text{EUreg}} = \beta_{\text{Stand}\times\text{EUgro}} = \beta_{\text{Inf}\times\text{EUgro}} = 0$, i.e. the EU regime does not moderate the content–pricing relationship. F-statistics from Panel C regressions (content × EU interaction, full controls, year and industry fixed effects). $df_1 = 4$ restrictions throughout; $df_2 = N - k$ varies by section and dependent variable. p-values use the $F(4, df_2)$ distribution. Two-way industry × year clustered standard errors. Significance: $^{***}\,p < 0.01$, $^{**}\,p < 0.05$, $^{*}\,p < 0.10$.

Table 22: PCA Robustness — Summary Statistics of Content Variables

	Baseline 3,496 dims	PCA-40 $k = 40$	PCA-80 $k = 80$	PCA-125 $k = 125$
Standard content — mean	1.0007	0.7086	0.6535	0.6247
Standard content — s.d.	0.1525	0.9882	0.9214	0.8955
Informative content — mean	0.5849	3.8195	5.3648	6.5644
Informative content — s.d.	0.1070	0.6021	0.7951	0.9680
N	263	263	263	263

Notes: Mean and standard deviation of standard and informative content for the total prospectus document across four specifications. Baseline uses the full 3,496-dimensional L^1-normalized word vectors; PCA-k compresses to k principal components (L^2-normalized) before the Hanley and Hoberg (2010) decomposition, retaining 80%, 90%, and 95% of total variance for $k = 40$, 80, and 125 respectively. Content variables are in original (unstandardized) units. N refers to IPOs for which the decomposition is feasible.

Table 23: PCA Robustness — Content and IPO Pricing

	Panel A: Content + controls				*Panel C: EU interaction*			
	Baseline 3,496 dims	PCA-40 $k = 40$	PCA-80 $k = 80$	PCA-125 $k = 125$	Baseline 3,496 dims	PCA-40 $k = 40$	PCA-80 $k = 80$	PCA-125 $k = 125$
Dep. var.: ΔP								
Standard (z)	−0.0473	−0.0384	−0.0424	−0.0410	—	—	—	—
	(−1.30)	(−1.01)	(−1.16)	(−1.12)				
Informative (z)	−0.0050	−0.0563	−0.0895*	−0.0970**	—	—	—	—
	(−0.10)	(−1.41)	(−1.83)	(−1.98)				
Stand (z) × EU Reg.	—	—	—	—	+0.0060	+0.0568	+0.0289	+0.0033
					(+0.08)	(+0.65)	(+0.38)	(+0.04)
Inf (z) × EU Reg.	—	—	—	—	+0.0866	+0.2040	+0.1869	+0.1772*
					(+0.70)	(+1.63)	(+1.58)	(+1.70)
Stand (z) × EU Growth	—	—	—	—	−0.2131	−0.0154	−0.0043	+0.0004
					(−1.58)	(−0.14)	(−0.04)	(+0.00)
Inf (z) × EU Growth	—	—	—	—	+0.3242	−0.0323	−0.0485	−0.0835
					(+1.26)	(−0.42)	(−0.68)	(−1.05)
N	257	257	257	257	257	257	257	257
R^2	0.128	0.131	0.139	0.142	0.141	0.147	0.153	0.155
Dep. var.: $\lvert\Delta P\rvert$								
Standard (z)	−0.0281	−0.0507	−0.0514	−0.0489	—	—	—	—
	(−0.81)	(−1.34)	(−1.41)	(−1.34)				
Informative (z)	+0.0232	−0.0686*	−0.1007**	−0.1087**	—	—	—	—
	(+0.44)	(−1.73)	(−2.09)	(−2.24)				
Stand (z) × EU Reg.	—	—	—	—	−0.0034	+0.0283	+0.0036	−0.0212
					(−0.05)	(+0.32)	(+0.05)	(−0.28)
Inf (z) × EU Reg.	—	—	—	—	+0.1224	+0.1863	+0.1621	+0.1502
					(+0.98)	(+1.50)	(+1.37)	(+1.43)
Stand (z) × EU Growth	—	—	—	—	−0.2395*	−0.0224	−0.0064	−0.0006
					(−1.85)	(−0.21)	(−0.06)	(−0.01)
Inf (z) × EU Growth	—	—	—	—	+0.4296*	−0.0311	−0.0486	−0.0925
					(+1.77)	(−0.41)	(−0.70)	(−1.21)
N	257	257	257	257	257	257	257	257
R^2	0.121	0.130	0.139	0.143	0.148	0.148	0.153	0.155

Notes: OLS regressions (Equation 9) of IPO pricing variables on standardized content measures, with two-way industry × year clustered standard errors. Panel A reports the baseline content–pricing relationship with full firm controls and year and industry fixed effects. Panel C adds four content × EU regime interaction terms. Content variables are z-scored within each specification so coefficients are comparable across compression levels. Baseline uses the full 3,496-dimensional word vectors; PCA-k compresses to k principal components before the Hanley and Hoberg (2010) decomposition, retaining 80%, 90%, and 95% of total variance for $k = 40$, 80, and 125 respectively. Controls: log firm age, log filing size, tech dummy, PE dummy, regulated market dummy, lead underwriter market share, OMX30 30-day return, log # IPOs (180d). t-statistics in parentheses. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table 24: Robustness Check — Standard-Error Specification

Cell	Variable	Section	HC1	Year	Ind×Year	Ind×Yr×UW
C1	Informative	Tot	+0.187 (+0.80)	+0.187 (+0.91)	+0.187 (+0.79)	+0.187 (+0.79)
C1	Standard	Tot	+0.117 (+0.72)	+0.117 (+1.62)	+0.117 (+0.77)	+0.117 (+0.77)
C2	Informative	Tot	+0.221 (+0.92)	+0.221 (+1.06)	+0.221 (+0.82)	+0.221 (+0.82)
C2	Standard	Tot	+0.226 (+0.74)	+0.226 (+1.43)	+0.226 (+0.84)	+0.226 (+0.84)
C3	Informative	Risk	−0.145 (−1.00)	−0.145 (−0.73)	−0.145 (−1.21)	−0.145 (−1.21)
C3	Standard	Risk	+0.156 (+1.42)	+0.156 (+1.60)	+0.156 (+1.08)	+0.156 (+1.08)
C4	Informative	Tot	+0.092 (+0.31)	+0.092 (+0.25)	+0.092 (+0.39)	+0.092 (+0.39)
C4	Standard	Tot	−0.231 (−1.19)	−0.231 (−1.07)	−0.231 (−0.78)	−0.231 (−0.78)
C5	Inf × EU Reg.	Risk	+0.071 (+0.26)	+0.071 (+0.25)	+0.071 (+0.28)	+0.071 (+0.28)
C5	Informative	Risk	−0.196 (−1.21)	−0.196 (−1.02)	−0.196 (−1.56)	−0.196 (−1.56)
C5	Standard	Risk	−0.029 (−0.31)	−0.029 (−0.30)	−0.029 (−0.24)	−0.029 (−0.24)
C5	Stand × EU Reg.	Risk	−0.204 (−1.22)	−0.204 (−1.18)	−0.204 (−0.78)	−0.204 (−0.78)
C6	EU Growth	Tot	+0.002 (+0.06)	+0.002 (+0.10)	+0.002 (+0.05)	+0.002 (+0.05)
C6	EU Reg.	Tot	+0.039 (+0.76)	+0.039** (+2.11)	+0.039 (+1.06)	+0.039 (+1.06)

Notes: OLS coefficients with t-statistics in parentheses under four cluster-robust covariance specifications: HC1 (heteroskedasticity-robust, no clustering), one-way year clustering, two-way industry × year clustering (baseline), and three-way industry × year × underwriter clustering. Cells C1–C3 are IR pricing specifications; C4–C5 are ΔP pricing specifications; C6 stress-tests the EU Regulation and EU Growth coefficients on total standard and informative content (Table 4). Cell definitions follow Appendix C. The joint Wald F-test of EU-regime moderation is stress-tested separately in Table 11. Significance: $^{***}\,p < 0.01$, $^{**}\,p < 0.05$, $^{*}\,p < 0.10$.

Table 25: Wald F-Robustness — Joint EU-Regime Moderation Across Tests

Tests	Section	F	df_1	df_2	p	N
Baseline (= Table 7)	Total	5.319	4	194	0.0004***	251
	Summary	3.535	4	177	0.0084***	230
	Risk	3.621	4	193	0.0072***	249
cluster = HC1	Total	1.112	4	194	0.3520	251
	Summary	2.978	4	177	0.0207**	230
	Risk	3.599	4	193	0.0074***	249
cluster = Year	Total	11.189	4	194	0.0000***	251
	Summary	11.205	4	177	0.0000***	230
	Risk	6.434	4	193	0.0001***	249
cluster = Ind×Yr×UW	Total	5.319	4	194	0.0004***	251
	Summary	3.535	4	177	0.0084***	230
	Risk	3.621	4	193	0.0072***	249
regulated market	Total	2.931	4	146	0.0229**	202
	Summary	0.813	4	134	0.5190	186
	Risk	3.752	4	145	0.0062***	200
MTF	Total	5.858	4	8	0.0167**	49
	Risk	4.548	4	8	0.0329**	49
excl. 2021	Total	2.211	4	124	0.0716*	179
	Summary	4.330	4	110	0.0027***	161
	Risk	0.867	4	123	0.4858	177

Notes: Joint Wald F-test of H_0: all four content × EU regime interaction coefficients equal zero, i.e. the EU regime does not moderate the content–pricing relationship for IR. All specifications use full controls and year and industry fixed effects. Cluster specifications: HC1 = heteroskedasticity-robust, no clustering; Year = one-way year clustering; Ind×Yr×UW = three-way clustering with PSD-fallback to two-way. Venue split: Regulated = Nasdaq Stockholm; MTF = First North, Spotlight, and NGM (Summary omitted for MTF due to insufficient observations). Excl. 2021 drops the IPO boom year. Significance: $^{***}\,p < 0.01$, $^{**}\,p < 0.05$, $^{*}\,p < 0.10$.

Table 26: Robustness Check — Pre/Post-EU Regulation Sub-Period Split (Full)

Specification	Variable	Period	Coef.	t	N
Total IR (full)	Standard	All	+0.226	+0.84	251
		Pre	+0.088	+0.35	113
		Post	+0.276	+0.40	138
	Informative	All	+0.221	+0.82	251
		Pre	+0.478**	+2.27	113
		Post	+0.086	+0.17	138
Risk IR (full)	Standard	All	+0.156	+1.08	249
		Pre	+0.278**	+2.36	111
		Post	+0.097	+0.35	138
	Informative	All	−0.145	−1.21	249
		Pre	−0.101	−0.52	111
		Post	−0.325	−1.46	138
Total ΔP (Panel A)	Standard	All	−0.231	−0.78	257
		Pre	−0.207	−0.54	117
		Post	−0.537	−1.21	140
	Informative	All	+0.092	+0.39	257
		Pre	−0.220	−0.33	117
		Post	+0.513	+1.20	140
Risk ΔP (Panel A)	Standard	All	−0.088	−0.55	255
		Pre	−0.054	−0.39	115
		Post	−0.193	−0.81	140
	Informative	All	−0.191***	−2.99	255
		Pre	−0.018	−0.14	115
		Post	−0.217	−0.84	140
Chow-style F-test (H_0: content coefficients identical pre/post):					
Total IR (full)	—	—	$F = 0.144$	$df = (2, 195)$	$p = 0.8656$
Risk IR (full)	—	—	$F = 0.830$	$df = (2, 194)$	$p = 0.4375$
Total ΔP (Panel A)	—	—	$F = 0.650$	$df = (2, 203)$	$p = 0.5230$
Risk ΔP (Panel A)	—	—	$F = 0.133$	$df = (2, 202)$	$p = 0.8758$

Notes: OLS regressions (Equation 8) with full controls, year and industry fixed effects, and two-way industry × year clustered standard errors. 'Pre' = filings before 21 July 2019; 'Post' = on or after. The Chow-style F-test pools both periods with content × period interactions and tests their joint equality to zero. It tests the same family of restrictions as the joint Wald F in Table 11, but with substantially less power as the effective sample for each sub-period is roughly half the pooled N. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table 27: Robustness Check — Bookbuilding Sub-Sample ($\Delta P \neq 0$, Full)

Section	Dep. var.	Specification	Variable	Full sample	BB only
Panel A: Baseline					
Risk	ΔP	Panel A	Informative	-0.191^{***} ($n = 255$)	$+0.002$ ($n = 49$)
Risk	ΔP	Panel A	Standard	-0.088 ($n = 255$)	-0.917 ($n = 49$)
Risk	$\lvert\Delta P\rvert$	Panel A	Informative	-0.133^{**} ($n = 255$)	$+0.550$ ($n = 49$)
Risk	$\lvert\Delta P\rvert$	Panel A	Standard	-0.038 ($n = 255$)	$+0.189$ ($n = 49$)
Sum	ΔP	Panel A	Informative	$+0.004$ ($n = 236$)	$+0.188$ ($n = 47$)
Sum	ΔP	Panel A	Standard	-0.055 ($n = 236$)	-0.069 ($n = 47$)
Sum	$\lvert\Delta P\rvert$	Panel A	Informative	$+0.015$ ($n = 236$)	$+0.388$ ($n = 47$)
Sum	$\lvert\Delta P\rvert$	Panel A	Standard	-0.079 ($n = 236$)	-0.317 ($n = 47$)
Tot	ΔP	Panel A	Informative	$+0.092$ ($n = 257$)	$+0.954$ ($n = 50$)
Tot	ΔP	Panel A	Standard	-0.231 ($n = 257$)	-4.150 ($n = 50$)
Tot	$\lvert\Delta P\rvert$	Panel A	Informative	$+0.345$ ($n = 257$)	$+2.073$ ($n = 50$)
Tot	$\lvert\Delta P\rvert$	Panel A	Standard	-0.137 ($n = 257$)	-1.720 ($n = 50$)
Panel C: EU interaction terms					
Risk	ΔP	Panel C	Informative	-0.196 ($n = 255$)	-2.560 ($n = 49$)
Risk	ΔP	Panel C	Standard	-0.029 ($n = 255$)	-0.014 ($n = 49$)
Risk	ΔP	Panel C	Inf × EU Reg.	$+0.071$ ($n = 255$)	$+8.390^{*}$ ($n = 49$)
Risk	ΔP	Panel C	Stand × EU Reg.	-0.204 ($n = 255$)	-6.045 ($n = 49$)
Risk	$\lvert\Delta P\rvert$	Panel C	Informative	-0.239^{**} ($n = 255$)	-2.712^{*} ($n = 49$)
Risk	$\lvert\Delta P\rvert$	Panel C	Standard	-0.012 ($n = 255$)	-0.837 ($n = 49$)
Risk	$\lvert\Delta P\rvert$	Panel C	Inf × EU Reg.	$+0.141$ ($n = 255$)	$+6.593$ ($n = 49$)
Risk	$\lvert\Delta P\rvert$	Panel C	Stand × EU Reg.	-0.188 ($n = 255$)	-3.321 ($n = 49$)
Tot	ΔP	Panel C	Informative	-0.236 ($n = 257$)	$+0.247$ ($n = 50$)
Tot	ΔP	Panel C	Standard	-0.044 ($n = 257$)	-1.409 ($n = 50$)
Tot	ΔP	Panel C	Inf × EU Reg.	$+0.365$ ($n = 257$)	$+6.876$ ($n = 50$)
Tot	ΔP	Panel C	Stand × EU Reg.	-0.226 ($n = 257$)	-2.774 ($n = 50$)
Tot	$\lvert\Delta P\rvert$	Panel C	Informative	-0.285 ($n = 257$)	-0.082 ($n = 50$)
Tot	$\lvert\Delta P\rvert$	Panel C	Standard	$+0.061$ ($n = 257$)	-0.815 ($n = 50$)
Tot	$\lvert\Delta P\rvert$	Panel C	Inf × EU Reg.	$+0.630$ ($n = 257$)	$+17.702$ ($n = 50$)
Tot	$\lvert\Delta P\rvert$	Panel C	Stand × EU Reg.	-0.286 ($n = 257$)	-0.063 ($n = 50$)

Notes: OLS coefficients (Equation 9) with significance stars and sample size for the full pricing sample versus the bookbuilding sub-sample ($\Delta P \neq 0$, $N \approx 49$–50). Full sample uses two-way industry × year clustered standard errors; bookbuilding sub-sample uses HC1 due to insufficient clustering variation. Panel A = content and controls only; Panel C adds content × EU regime interaction terms. IR results are not stress-tested as the IR sample is unrestricted by the $\Delta P = 0$ institutional feature. The check supports the decision to lead the pricing analysis with IR rather than ΔP. Significance: $^{***}\,p < 0.01$, $^{**}\,p < 0.05$, $^{*}\,p < 0.10$.

Table 28: Robustness Check — Decomposition Window

Specification	Variable	60d	90d (baseline)	120d	180d
Total IR (full)	Informative	+0.355 (+1.18)	+0.221 (+0.82)	+0.243 (+0.91)	+0.225 (+0.87)
	Standard	+0.257 (+0.89)	+0.226 (+0.84)	+0.224 (+0.78)	+0.211 (+0.71)
Total ΔP (Panel A)	Informative	+0.124 (+0.54)	+0.092 (+0.39)	+0.019 (+0.08)	+0.187 (+0.83)
	Standard	−0.211 (−0.74)	−0.231 (−0.78)	−0.282 (−0.88)	−0.260 (−0.88)
Risk IR (full)	Informative	−0.077 (−0.69)	−0.145 (−1.21)	−0.121 (−0.97)	−0.129 (−0.99)
	Standard	+0.148 (+1.08)	+0.156 (+1.08)	+0.146 (+1.06)	+0.145 (+1.02)
Risk IR (Panel C)	Standard	+0.251*** (+3.05)	+0.251*** (+3.11)	+0.258*** (+3.19)	+0.329*** (+3.99)
	Informative	+0.020 (+0.15)	−0.021 (−0.15)	+0.017 (+0.12)	+0.034 (+0.22)
	Stand × EU Reg.	−0.552*** (−3.12)	−0.537*** (−2.74)	−0.531*** (−3.29)	−0.543*** (−3.56)
	Inf × EU Reg.	−0.134 (−0.76)	−0.143 (−0.82)	−0.161 (−0.81)	−0.234 (−0.88)
	Stand × EU Gr.	+1.131*** (+2.86)	+1.100*** (+2.99)	+1.151*** (+2.85)	+0.916** (+2.31)
	Inf × EU Gr.	−0.662** (−2.01)	−0.711** (−1.99)	−0.766* (−1.89)	−0.625 (−1.61)
Pairwise correlations of Total informative content across windows:					
60d vs. 90d		$\rho = 0.974$, $N = 262$			
60d vs. 120d		$\rho = 0.958$, $N = 262$			
60d vs. 180d		$\rho = 0.926$, $N = 262$			
90d vs. 120d		$\rho = 0.984$, $N = 263$			
90d vs. 180d		$\rho = 0.954$, $N = 263$			
120d vs. 180d		$\rho = 0.973$, $N = 263$			

Notes: OLS coefficients with t-statistics in parentheses, full controls, year and industry fixed effects, and two-way industry × year clustered standard errors (Equations 8 and 9). The Hanley and Hoberg (2010) decomposition is recomputed for each window length (60, 90 (baseline), 120, and 180 days) and the headline regressions re-estimated on the window-specific content variables. Pairwise correlations of total informative content confirm near-identical content measures across windows ($\rho > 0.93$ in all cases), supporting the robustness of results to the choice of decomposition window. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table 29: Robustness Check — Underwriter Fixed Effects

Specification	Variable	None	≥5 deals	≥10 deals	Full (≥2 deals)
Total IR (full)	Standard	+0.226 (+0.84)	+0.256 (+0.85)	+0.199 (+0.67)	+0.456*** (+3.25)
	Informative	+0.221 (+0.82)	+0.151 (+0.64)	+0.077 (+0.34)	+0.004 (+0.01)
Risk IR (full)	Standard	+0.156 (+1.08)	+0.108 (+0.81)	+0.112 (+0.76)	+0.215 (+1.38)
	Informative	−0.145 (−1.21)	−0.136 (−0.96)	−0.166 (−1.15)	−0.105 (−0.78)
Total ΔP (Panel A)	Standard	−0.231 (−0.78)	−0.135 (−0.67)	−0.053 (−0.27)	−0.150 (−0.56)
	Informative	+0.092 (+0.39)	−0.005 (−0.03)	−0.071 (−0.62)	−0.019 (−0.10)
Risk ΔP (Panel A)	Standard	−0.088 (−0.55)	−0.091 (−0.60)	−0.047 (−0.33)	−0.088 (−0.45)
	Informative	−0.191*** (−2.99)	−0.089 (−1.08)	−0.135** (−2.45)	−0.123 (−1.03)

Notes: OLS coefficients (Equations 8 and 9) with t-statistics in parentheses, full controls, year and industry fixed effects, and two-way industry × year clustered standard errors. UW FE columns add dummies for underwriters appearing in $\geq N$ IPOs; remaining underwriters are pooled into an 'Other UW' baseline. The Full (≥2 deals) column is the maximal specification; the significant Standard coefficient for Total IR in this column should be interpreted with caution as near-full saturation may absorb economically relevant variation. The continuous lead-UW market-share control is included in all specifications. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table 30: Robustness Check — Winsorisation Level

Specification	Variable	None	1% (baseline)	2.5%	5%
Total IR (full)	Standard	+0.224 (+0.84)	+0.226 (+0.84)	+0.454* (+1.77)	+0.385** (+2.31)
	Informative	+0.228 (+0.84)	+0.221 (+0.82)	+0.412 (+1.03)	+0.266 (+0.95)
Total ΔP (Panel A)	Standard	−0.226 (−0.59)	−0.231 (−0.78)	−0.091 (−0.46)	+0.000 (+0.02)
	Informative	−0.141 (−0.34)	+0.092 (+0.39)	−0.010 (−0.05)	−0.051** (−2.44)
Risk IR (full)	Standard	+0.156 (+1.08)	+0.156 (+1.08)	+0.118 (+0.83)	+0.101 (+0.94)
	Informative	−0.143 (−1.16)	−0.145 (−1.21)	+0.179 (+0.58)	+0.078 (+0.48)
Risk ΔP (Panel A)	Standard	−0.060 (−0.26)	−0.088 (−0.55)	−0.040 (−0.49)	−0.008 (−0.68)
	Informative	−0.367** (−2.09)	−0.191*** (−2.99)	−0.148*** (−2.76)	−0.016 (−1.55)

Notes: OLS coefficients (Equations 8 and 9) with t-statistics in parentheses. Pricing variables are re-winsorised at four alternative cut-points: none (0%), 1% (baseline), 2.5%, and 5% on each tail. Two-way industry × year clustered standard errors. Full controls and year and industry fixed effects included throughout. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

H. AI Disclosure

AI tools, including Claude AI (Anthropic) and ChatGPT, were used in three capacities when writing this thesis. First, as a methodological aid – helping us understand and implement the text analysis and regression framework from Hanley and Hoberg (2010) as well as the methodology in Appendix D & F. Second, as a computation tool – assisting in the writing and debugging of Python code used for data collection, text processing and the statistical estimation. Third, as a language aid – reviewing and improving the clarity and grammar of the written text, while all arguments, interpretations and conclusions remain our own. All AI-assisted outputs were critically reviewed and verified before use.