What Do Public Firms Reveal About Firms Going Public?

Mingjun Sun*

November 21, 2026

Click here for the latest version

Abstract

I study what economically similar public firms reveal about the financing, pricing, and risk of firms going public. Using a pretrained embedding model, I identify each issuer's connected firms by matching business descriptions in IPO registration statements to those in seasoned firms' 10-Ks. Connected stock returns before listing, including those realized during bookbuilding, predict first-day IPO returns. Offer prices reflect only part of this signal, with further adjustment during first-day trading. The total price response to connected firms' earnings-announcement returns is larger when issuers' initial filings contain more uncertain language. Issuers have higher leverage at and after listing when their connected firms are more highly levered. Among issuers whose pre-IPO leverage exceeds connected leverage, a larger gap predicts a greater reduction in debt around the offering. After listing, issuer returns comove with connected returns beyond industry movements. Connected firms' pre-offer total and idiosyncratic volatility predict issuer total volatility throughout the first trading year.

*Tepper School of Business, Carnegie Mellon University, Pittsburgh, PA 15213, USA. Email: mingjuns@andrew.cmu.edu. I am indebted to my advisors, Burton Hollifield, Bryan Routledge, Patrick Blonien and Miguel Oliveira, for their invaluable guidance and support. For helpful comments, suggestions, and discussions, I thank Chester Spatt and seminar participants at Carnegie Mellon University.

1 Introduction

Before a firm goes public, investors and underwriters have its accounts and prospectus, but no history of how its stock responds to economic news. Public firms with related businesses provide another source of information. This paper asks what economically similar public firms reveal about an issuer's financing, pricing, and risk.

Firms operating related businesses may face common changes in demand, competition, and technology. Their stock returns can therefore convey information about an issuer's prospects before its own shares trade. Yet it is not obvious how much of this public information is reflected in the IPO offer price rather than in first-day trading. Similar business exposures may also make public firms' leverage and volatility useful benchmarks for the issuer's financing and equity risk. The IPO creates an opportunity to examine these relationships around a financing transition: access to public equity changes, while the issuer's underlying products, customers, and competitive environment need not change at the same time.

Identifying the relevant public firms is difficult. Industry classifications place businesses into common categories, yet firms within a category can differ substantially in their products, customers, and sources of revenue. Conversely, firms assigned to different industries may conduct closely related activities. Firm characteristics such as size, profitability, and asset tangibility capture important differences but do not fully describe what a firm does. The choice of comparables matters for relative valuation (Alford, 1992; Bhojraj and Lee, 2002; Liu et al., 2002), including the valuation of IPOs (Kim and Ritter, 1999; Purnanandam and Swaminathan, 2004). For a new issuer, selecting comparables from observed return correlations is not yet possible.

Business descriptions offer a way to identify these firms before the issuer begins trading. They describe activities that can connect firms within and across industry boundaries (Hoberg and Phillips, 2010, 2016, 2025). I match business descriptions in issuers' S-1/F-1 registration statements to those in seasoned public firms' 10-K reports using a pretrained language embedding model. The model represents the content of each description numerically, allowing firms to be ranked by business similarity. Company names and other identifiers are masked before matching. Each issuer's ten closest public firms form its connected set. The matches use filings available by the initial range filing, and the connected set is held fixed across the analyses. The resulting network covers 4,207 U.S. IPOs from 1996 through 2025. A common set of firms therefore connects the information available before listing to several aspects of

the issuer's transition to public markets.

Connected firms' returns during bookbuilding predict first-day IPO returns even after conditioning on the offer price revision. A one-standard-deviation increase in connected bookbuilding returns is associated with a first-day closing price approximately 3.44% higher relative to the offer price, controlling for the revision, pre-range returns, market and industry returns, and issuer characteristics. Connected returns during bookbuilding also predict upward offer price revisions. The conditional first-day relation remains when I separately add returns on firms selected using finer industry classifications, size, or accounting characteristics.

Connected returns measured over a broader pre-listing window also predict first-day returns. Over the six months before the offering, an increase of one standard deviation in connected returns, 35.70 percentage points, is associated with a first-day return higher by 10.06 percentage points, compared with a sample mean of 25.2%. This estimate controls for contemporaneous market and industry returns and issuer characteristics, without conditioning on the offer price revision. In a separate decomposition of log price changes from the initial range midpoint to the first-day close, about 40% of the estimated response to six-month connected returns appears in the offer price revision, with the remainder appearing during initial trading. These results are consistent with partial offer price adjustment to public information about the issuer's business. A higher offer price increases the proceeds received for the shares sold, whereas a higher first-day return accrues to investors who receive shares at that price.

The response to connected firms' earnings-announcement news also varies with the issuer's valuation uncertainty. For a one-standard-deviation increase in connected announcement news, the estimated percentage change in price from the initial range midpoint to the first-day close is approximately 2.27 percentage points larger when uncertainty in the initial S-1/F-1 is one standard deviation higher. This pattern is consistent with related public firms providing more useful valuation information when the issuer's own prospects are harder to assess.

The leverage of connected firms also provides a benchmark for financing changes around the IPO. Among issuers whose pre-IPO leverage exceeds connected leverage, a connected leverage ratio that is one standard deviation (11.80 percentage points) lower is associated with an additional debt reduction of 2.98% of pre-IPO assets, holding initial leverage and other controls fixed. The estimate is 2.73% after controlling for leverage predicted from firm characteristics. Because debt changes are scaled by assets measured before the offering, the

result captures a reduction in debt outstanding rather than simply a lower debt ratio caused by the addition of new equity and assets.

The adjustment is asymmetric. A larger gap above connected leverage is associated with a greater debt reduction around listing, whereas the corresponding relation for issuers starting below connected leverage is close to zero. Access to new equity may facilitate debt reduction, while issuers below connected leverage may defer borrowing. Issuers farther below connected leverage undertake more gross long-term borrowing during the following year, conditional on initial leverage and conventional financing benchmarks.

Connected leverage also predicts differences in capital structure after listing. An increase of 11.80 percentage points in connected book leverage is associated with issuer book leverage that is 2.04 percentage points higher one year after listing and 2.55 percentage points higher after two years, conditional on leverage before the IPO, industry leverage, and issuer characteristics.

Once issuers begin trading, their returns provide a test of whether the firms selected from business descriptions share common economic exposures with the issuers. An increase of one standard deviation in connected monthly returns is associated with a contemporaneous issuer return that is 3.58 percentage points higher, after controlling for industry returns and issuer and calendar-month fixed effects. The relation also holds over months 13–36 after listing. At the firm-pair level, greater pre-IPO business similarity predicts higher factor- and industry-adjusted return correlation and greater sales-growth comovement, while the profitability evidence is mixed. Since issuer returns play no role in forming the connected set, these patterns support the economic relevance of the pre-IPO matches.

The connections also extend beyond firms explicitly named as competitors. Among issuers with usable competitor disclosures, public firms selected from business descriptions but absent from those lists explain issuer returns beyond the named firms. Named competitors, in turn, explain returns beyond the text matches.

Connected firms' pre-offer total volatility also predicts issuer total volatility throughout the first year of trading, conditional on industry total volatility and issuer characteristics. An increase of one standard deviation in log connected volatility corresponds to approximately 48% higher connected volatility before the offer. At the respective sample medians, this difference is associated with 7.23 percentage points higher annualized issuer volatility over trading days 2–20 and 4.67 percentage points higher volatility in the second half of the first year. The relation is positive across all four windows examined. Connected firms' pre-offer

idiosyncratic volatility predicts issuer total volatility in separate specifications as well. These findings extend research on valuation difficulty at IPOs (Lowry et al., 2010), prospectus language and volatility (Loughran and McDonald, 2013), common variation in idiosyncratic volatility (Herskovic et al., 2016), and learning about risk from comparable firms (Kim et al., 2024).

Comparisons with alternative comparable sets distinguish the role of business descriptions from the choice of text model. Following Lee et al. (2015), I compare explanatory power using identical observations within each pairwise comparison. The connected measures explain more variation than conventional industry classifications in leverage, pricing, monthly returns, and volatility. Embeddings improve on word-frequency matching in explaining monthly returns and initial volatility, but the differences in explanatory power for IPO pricing are not statistically significant. Matching on financial characteristics explains slightly more variation in initial issuer volatility than embedding-based matching.

The paper contributes primarily to research on comparable firms and IPO price formation. Relative valuation studies examine how comparable firms help assess an issuer's value (Kim and Ritter, 1999; Purnanandam and Swaminathan, 2004), while research on bookbuilding examines information elicited from investors and the relation between price revisions and initial returns (Hanley, 1993; Cornelli and Goldreich, 2001, 2003). Lowry and Schwert (2004) and Edelen and Kadlec (2005) provide the closest evidence on adjustment to public information. Lowry and Schwert find a small incremental relation between public information and initial returns after controlling for the offer price revision. Edelen and Kadlec use industry-comparable returns and explain partial adjustment through the issuer's tradeoff between proceeds and the probability of completing the offering. Using issuer business descriptions to select comparables within and across industries, I find that their bookbuilding returns predict first-day returns after conditioning on the offer price revision and conventional comparable returns.

The financing evidence relates to work on industry conditions and corporate financial policy (MacKay and Phillips, 2005; Leary and Roberts, 2014; Grieser et al., 2022), and on the persistence and adjustment of capital structure. Lemmon et al. (2008) document persistent leverage differences that are already present before firms go public; DeAngelo and Roll (2015) show substantial movement in individual leverage ratios and firms' relative positions over time. Studies of IPO financing and leverage adjustment examine how financing opportunities

and initial positions shape subsequent changes (Pagano et al., 1998; Alti, 2006; Leary and Roberts, 2005; Faulkender et al., 2012). I study the IPO as a point at which firms can revise financing positions inherited from private ownership. Connected leverage helps explain the asymmetric debt changes around the offering and subsequent leverage differences, conditional on the issuer's starting capital structure.

The post-IPO comovement evidence relates to research on the economic links among firms. Customer and supplier ties, competition, and production complementarities help explain comovement and return predictability (Cohen and Frazzini, 2008; Menzly and Ozbas, 2010; Eisdorfer et al., 2022; Lee et al., 2024). Related firms' stock prices can also inform corporate investment (Foucault and Frésard, 2014), while industry conditions matter for interpreting the effects of IPOs on rivals (Spiegel and Tookes, 2020). Here, the links are identified before the issuer has a public return history, and the subsequent comovement provides evidence about common exposures.

The matching approach builds on research that uses corporate disclosure to identify economic relationships among firms. Hoberg and Phillips (2010, 2016, 2018, 2025) develop and apply text-based measures of product market relationships, and Breitung and Müller (2025) construct business networks using language embeddings. Research on IPO prospectuses and regulatory review examines information production, disclosure, litigation risk, and pricing (Hanley and Hoberg, 2010, 2012; Loughran and McDonald, 2013; Lowry et al., 2020). I extend these business comparisons to issuers without a public trading history by matching business descriptions in registration statements to those in seasoned firms' annual reports. Related machine-learning research uses models to forecast stock returns (Gu et al., 2020), language representations to measure corporate culture (Li et al., 2021), generative AI to extract stakeholders' assessments of corporate culture (Li et al., 2026), and embeddings to measure public perceptions of finance (Jha et al., 2026).

The remainder of the paper proceeds as follows. Section 2 describes the data and IPO sample. Section 3 constructs the connected sets and compares them with alternative measures of relatedness. Section 4 examines connected returns, offer price revisions, and first-day returns. Section 5 studies issuer leverage and debt changes around the offering. Section 6 examines return comovement after listing and the complementary information in named competitors. Section 7 studies connected volatility and issuer risk. Section 8 concludes. The appendices provide additional results and details on data construction and inference.

2 Data and IPO Sample

2.1 IPO Sample

The initial sample consists of 4,619 IPOs completed in U.S. public markets from 1996 through 2025. I identify offerings using Jay Ritter's IPO data, SDC, IPO Scoop, and the Nasdaq IPO calendar, and link issuers to CRSP and SEC registration statements. The sample includes ordinary common stock listed on the NYSE, AMEX, or Nasdaq with an offer price of at least \$5. It excludes ADRs, unit offerings, SPACs, closed-end funds, REITs, banks and savings institutions, natural resource limited partnerships, and best efforts offerings with proceeds below \$10 million.

Identifying public comparables requires business descriptions for both the issuer and eligible public firms. Under the filing rules in Section 3, 4,227 issuers have a substantive Business section and 4,207 can be matched to public comparables. These 4,207 IPOs form the sample for the text-based analysis. Coverage is lowest in 1996, when the SEC's electronic filing mandate was taking effect. Appendix E reports annual coverage and shows that the main estimates remain similar when 1996 is excluded.

2.2 Offering Characteristics and Financial Data

SDC provides offer prices, shares offered, and gross proceeds, supplemented first by Nasdaq and then, for offer prices, by IPO Scoop. I use reported gross proceeds in Table 1 and the leverage regressions, and offer price times shares offered in the return and pricing regressions. Ritter's data provide founding years, venture capital and growth capital backing, and the Internet indicator. Firm age is the offer year minus the founding year; the backing indicator equals one for either venture capital or growth capital backing.

The pricing tests also require the initial filing range. I use the earliest valid range in SEC filings before the offer, supplemented by SDC when an SEC range is unavailable. Tests that separate information arriving before and during bookbuilding require an SEC range with an observed filing date, so that connected stock returns can be measured relative to the range filing.

Daily and monthly stock prices, returns, and market equity come from CRSP. Annual and quarterly accounting data for issuers and their public comparables come from Compustat, linked through the CRSP/Compustat linking table. The Fama–French data supply the market

return and the factors used to estimate idiosyncratic volatility. The accounting measures and their timing are defined in Section 5; post-IPO returns and volatility are defined in Sections 6 and 7.

2.3 Summary Statistics

Table 1 summarizes the IPO sample and its public comparables. The first-day pricing sample contains 3,954 IPOs, and the book and market leverage samples contain 3,728 and 3,726 IPOs, respectively.

The median issuer is nine years old and raises $79.80 million, and just over half of the issuers have venture capital or growth capital backing. First-day returns average 25.16%, with a median of 11.93%. Financing positions at listing vary widely: issuer book leverage ranges from 0.09% at the 25th percentile to 22.06% at the 75th percentile. The median change in debt around the offering is −1.81% of pre-IPO assets.

The interquartile range of connected book leverage is 8.51%–25.25%, and that of six-month pre-offer connected returns is −1.22%–28.65%. Issuers exhibit higher median idiosyncratic volatility after listing than their comparables exhibit before the offer: annualized volatility is 61.91% over issuer trading days 21–63, compared with 48.63% for connected firms over the six months before the offer.

3 Identifying Connected Firms

3.1 Identifying Connected Firms from Business Descriptions

I identify connected firms by comparing an issuer's business description with those of seasoned public firms. The matches use disclosures available before pricing to select firms with similar operations, products, customers, and competitive environments.

Business descriptions contain information about firms' product market positions and scope that conventional classifications can miss (Hoberg and Phillips, 2016, 2025). Recent work also uses language model embeddings of business descriptions to construct firm-level business networks (Breitung and Müller, 2025). I use this approach to identify public firms that are economically close to an issuer before the issuer itself begins trading.

For each issuer, I use disclosures available by the initial range filing. Issuer business descriptions come from the first substantive registration statement available by this date.

Table 1: Summary Statistics

	N	Mean	SD	Median	P25	P75
Panel A: IPO characteristics						
First-day return (%)	3,954	25.16	44.51	11.93	0.00	31.67
Proceeds ($ millions)	3,954	171.28	291.06	79.80	40.60	164.71
Offer price ($)	3,954	14.51	6.01	14.00	10.00	17.00
Firm age (years)	3,954	16.51	21.26	9.00	5.00	17.00
VC- or growth-capital-backed (%)	3,954	52.18	49.96	100.00	0.00	100.00
Internet firm (%)	3,954	15.58	36.27	0.00	0.00	0.00
Log offer price revision (×100)	3,954	-2.19	23.04	0.00	-14.31	10.54
Panel B: Issuer fundamentals						
Assets ($ millions)	3,728	515.17	1,220.09	128.91	57.01	348.15
Profitability (%)	3,728	-30.34	86.64	-1.63	-46.99	13.50
Tangibility (%)	3,728	19.78	20.90	11.97	5.31	26.12
Book leverage (%)	3,728	15.09	21.98	3.92	0.09	22.06
Market leverage (%)	3,726	9.34	16.10	1.12	0.02	10.95
Panel C: Connected firm measures						
Connected firm book leverage (%)	3,728	17.89	11.80	15.58	8.51	25.25
Connected firm market leverage (%)	3,726	12.63	11.21	9.04	3.90	18.67
Six-month connected return (%)	3,954	19.36	35.70	11.77	-1.22	28.65
Connected stock IVOL (%)	3,954	54.07	22.61	48.63	36.82	67.36
Average cosine similarity	3,954	0.84	0.05	0.84	0.81	0.87
Panel D: Main outcomes						
Debt change / pre-IPO assets (%)	3,699	-15.19	40.86	-1.81	-19.56	0.00
Post-IPO IVOL, days 21–63 (%)	3,951	69.40	35.10	61.91	44.43	85.67
Initial range width (%)	3,954	15.00	4.75	14.29	12.50	17.39
Absolute log revision (×100)	3,954	16.72	16.01	11.78	6.06	24.12
Market-adjusted BHAR, 12 months (%)	3,913	-11.04	72.91	-26.30	-62.48	20.70
Market-adjusted BHAR, 24 months (%)	3,861	-21.07	105.13	-49.78	-85.87	12.00
Market-adjusted BHAR, 36 months (%)	3,821	-27.36	131.06	-60.32	-105.78	7.78

Notes: Each row uses available observations within the fixed sample for the corresponding result family, and N counts distinct IPOs. Issuer assets, profitability, tangibility, debt change, and issuer and connected book leverage use the 3,728-IPO book leverage sample. Issuer and connected market leverage use the separate 3,726-IPO market leverage sample. All other rows use the 3,954-IPO first-day pricing sample. Missing values reduce the reported row counts within each family; the samples are not intersected across families. Continuous variables are winsorized at the 1st and 99th percentiles, as in the regressions. Indicators are expressed in percent. Firm age is the offer year minus the founding year for issuers with valid founding years. The first-day return is the return from the offer price to the CRSP close on the first public trading day. Price revision is $100 \ln(P^{offer}/P^{mid})$, where P^{mid} is the midpoint of the initial filing range; range width is $100(P^{high} - P^{low})/P^{mid}$. Issuer accounting and debt change measures exclude financial firms and utilities. Assets and leverage use the first quarter ending on or after the offer within 270 days; profitability and tangibility use the latest annual statement before the offer. Book leverage is debt/assets; market leverage is debt/(debt+market equity); profitability is operating income before depreciation/assets; tangibility is net property, plant and equipment/assets. Debt change compares quarterly debt before and after the offer, scaled by pre-IPO assets. Connected measures use equal weights among the top ten comparables with available data; connected returns begin six calendar months before the recorded offer date and end at the market close immediately before first public trading. Connected IVOL retains its six-month pre-offer estimation window. IVOL is annualized residual volatility from the Fama–French five-factor model plus momentum. BHAR is the issuer's buy-and-hold return after the first close less the market return over the stated horizon; Appendix E details its construction.

Public-firm descriptions come from prior-calendar-year 10-Ks in a Russell 3000-linked filing corpus. Eligible comparables must be trading at the information cutoff. Appendix E details filing selection, security links, and data availability.

A pretrained model may contain information learned after a historical IPO. Following Breitung and Müller (2025), I mask company names and other identifiers in both corpora before embedding. Models may still infer firm identity from masked text (Sarkar and Vafa, 2024).

I construct embeddings using OpenAI's text-embedding-3-small model with 1,536 dimensions and normalize each document vector to unit Euclidean norm. Let $e_i^S \in \mathbb{R}^{1536}$ denote the issuer vector and $e_{f,\tau}^K \in \mathbb{R}^{1536}$ the vector for firm f's public filing in calendar year τ. Their inner product is cosine similarity.

For IPO i in offer year $y(i)$, I rank eligible public firms by cosine similarity and retain the most similar filing for each firm. The set of the top K connected firms is

$$\mathcal{C}_i(K) \equiv \Big\{ f : \text{rank}_i(f) \leq K \Big\}, \qquad \text{rank}_i(f) \text{ orders firms by } e_i^{S\prime} e_{f,\, y(i)-1}^K.$$

The resulting sample comprises 4,207 IPOs, each matched to up to 30 public firms. The baseline analyses use the ten closest firms ($K = 10$); the appendix reports alternative set sizes. Connected sets are formed at the information cutoff and held fixed across tests, with each measure using the available data for those firms.

Weighting firms by pre-IPO cosine similarity leaves the estimates and statistical conclusions essentially unchanged (Appendix Table E.3). Similarity scores vary little within the selected sets, so the change from equal weighting is modest.

3.2 Illustrative Matches and Economic Validation

Table 2 reports the five closest connected firms for four illustrative IPOs selected for economic interpretability. Appendix Table E.2 reports eight illustrative IPOs from different periods and industries. The matches use masked descriptions; company names are restored from metadata for presentation. The firms are listed in descending similarity order.

Visa and MasterCard both operate payment networks, despite different recorded two-digit SIC codes. Synchrony's matches connect consumer credit with card networks and payment processing. Medline's matches span medical devices and healthcare distribution, while Conoco's share oil and gas production and related activities. The examples illustrate

how business descriptions identify relationships that extend across conventional industry boundaries.

Table 2: Illustrative Examples of Connected Firms

IPO firm	Year	Business	Closest connected firms	Business activities underlying the match
Medline	2025	Medical supplies and distribution	(1) Omnicell; (2) Insulet; (3) Pulmonx; (4) AngioDynamics; (5) Henry Schein	(1) Medical supplies and medication-management equipment; (2) Insulin-delivery devices; (3) Devices for emphysema treatment; (4) Vascular and cancer-treatment devices; (5) Medical and dental supplies distribution
Synchrony Financial	2014	Consumer credit and retail financing	(1) Mastercard; (2) Discover Financial Services; (3) Heartland Payment Systems; (4) American Express; (5) Higher One Holdings	(1) Electronic payment networks; (2, 4) Consumer credit cards; (3) Retail credit and merchant payment processing; (5) Student refund disbursement and payment services
Visa	2008	Electronic payments network	(1) MasterCard; (2) American Express; (3) Online Resources; (4) Net 1 UEPS Technologies; (5) Western Union	(1) Payment networks across SIC2 codes; (2) Card-payment networks; (3) Electronic payments and bill payment; (4) Smart-card payment processing; (5) Money transfer and consumer payment services
Conoco	1998	Oil and gas production and refining	(1) Chevron; (2) Occidental Petroleum; (3) Unocal; (4) Snyder Oil; (5) Mobil	(1) Oil and gas production, refining, and marketing; (2, 3, 4) Oil and gas exploration and production; (5) Oil and gas production, refining, and chemicals

Notes: The table lists the five public firms with the most similar business descriptions for each illustrative IPO. Firms are ordered by cosine similarity, and the numbers link each firm to its business activities. Matches use disclosures available before the offering. Appendix Table E.2 presents additional examples, and Appendix E describes selection.

Following Lee et al. (2015), Table 3 compares the explanatory power of embedding-based connections with that of comparable sets constructed from SIC classifications, observable firm characteristics, and TF-IDF similarity.[1] Each pairwise comparison uses identical observations and separate regressions of issuer outcomes on the corresponding comparable-set signals, with an intercept but no additional controls or fixed effects.

[1]The industry sets contain all eligible firms in the issuer's two-digit or four-digit SIC industry. Characteristic matching selects the ten closest firms across industries in standardized log assets, profitability, and tangibility. TF-IDF matching selects the ten closest firms by cosine similarity. Appendix E describes the construction of these benchmarks.

Embedding-based connections explain more variation than SIC benchmarks across the financing, pricing, comovement, and volatility outcomes. They also outperform characteristic matching for post-IPO book leverage, IPO pricing, and monthly comovement, although characteristic matching performs somewhat better for initial trading volatility. Relative to TF-IDF, embeddings provide significantly greater explanatory power for post-listing comovement and initial volatility, while the pricing differences are statistically insignificant. The following sections examine these relationships conditional on issuer characteristics.

4 Connected Stock Returns and IPO Pricing

Before an issuer begins trading, changes in its valuation cannot be observed directly from its own stock price. Returns on economically similar public firms provide an alternative signal about changes in demand, competitive conditions, and growth opportunities relevant to the issuer. During bookbuilding, this information may be reflected in revisions to the initial filing range, but the offer price and the first-day market price are determined at different stages of the IPO process. If offer prices do not fully incorporate information conveyed by connected firms, their returns during bookbuilding may remain associated with first-day returns even after controlling for the offer-price revision. I examine how connected returns relate to each pricing stage and whether these relationships extend beyond market, industry, and conventional comparable-firm returns.

The analysis extends research on comparable-firm valuation (Kim and Ritter, 1999), bookbuilding (Cornelli and Goldreich, 2001, 2003), and partial adjustment to public information (Hanley, 1993; Lowry and Schwert, 2004; Edelen and Kadlec, 2005) by using connected returns measured before the IPO.

4.1 Connected Stock Returns and First-Day Returns

The first-day return is

$$R_{i,0}^{\text{IPO}} = \frac{P_{i,0}^{\text{CRSP}}}{\text{OfferPrice}_i} - 1,$$

where $P_{i,0}^{\text{CRSP}}$ is the closing price on the issuer's first public trading day. I use the close on that date and retain the recorded offer date separately.

For each issuer, the connected return is the equal-weighted average of its connected firms' cumulative returns. The baseline uses the ten closest firms and a window beginning six

Table 3: Explanatory Power of Text-Based and Conventional Comparable Sets

Comparable set	Text R^2 (%)	Benchmark R^2 (%)	Difference (pp)	p-value
Panel A: Book leverage at +1 year				
SIC2	18.56	11.51	7.05^{***}	0.0005
SIC4	19.46	10.73	8.73^{***}	0.0005
Characteristics	18.74	4.89	13.85^{***}	0.0005
TF-IDF	18.61	16.82	1.79^{*}	0.0660
Panel B: Offer-price revision				
SIC2	7.64	4.97	2.67^{***}	0.0005
SIC4	7.12	4.05	3.08^{***}	0.0005
Characteristics	9.25	6.71	2.54^{***}	< 0.0100
TF-IDF	7.67	8.09	-0.42	0.4948
Panel C: Log first-day return				
SIC2	11.12	5.85	5.27^{***}	0.0005
SIC4	10.80	5.39	5.41^{***}	0.0005
Characteristics	12.12	8.19	3.93^{***}	0.0010
TF-IDF	11.11	10.23	0.89	0.3208
Panel D: Monthly comovement, months 1–36				
SIC2	2.60	2.01	0.59^{***}	< 0.0001
SIC4	2.63	1.91	0.72^{***}	< 0.0001
Characteristics	2.71	1.13	1.58^{***}	< 0.0001
TF-IDF	2.61	2.42	0.19^{**}	0.0448
Panel E: Log total volatility, days 2–20				
SIC2	30.97	22.01	8.96^{***}	0.0005
SIC4	30.45	21.38	9.07^{***}	0.0005
Characteristics	29.94	31.91	-1.97^{*}	0.0820
TF-IDF	30.87	28.55	2.32^{***}	0.0025

Notes: Text denotes embeddings. Entries are R^2 in percent; text-minus-benchmark differences are percentage points. Each row uses identical observations for regressions on the respective comparable-set signals, with an intercept and no additional controls or fixed effects. Panel A relates book leverage one year after listing to pre-IPO comparable leverage. Panels B and C use log offer-price revisions and log first-day returns, with pre-range and bookbuilding returns entering jointly; bookbuilding ends at $T-2$. Panels A–C and E use IPO cross sections; Panel D averages calendar-month cross-sectional R^2 equally across months, using issuers in their first 36 complete post-listing months. Panel E relates log issuer total volatility over days 2–20 to log comparable volatility over the six pre-offer months. Cross-sectional variables are winsorized at the 1st and 99th percentiles within each pairwise sample. Two-sided p-values use paired IPO bootstraps for Panels A–C and E and Newey–West inference for Panel D. Appendices E and F give benchmark construction and inference details. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

calendar months before the recorded offer date and ending at the market close immediately before first public trading. Connected firms must have valid returns on at least 60% of the trading days in the window. I construct industry and market returns over the same interval. The industry benchmark is the equal-weighted return of public firms in the issuer's two-digit SIC industry, excluding the issuer and, where possible, its connected firms. The market benchmark compounds the Fama–French daily market excess return plus the risk-free rate.

I estimate

$$R_{i,0}^{\text{IPO}} = \alpha + \beta\,\text{ConnRet}_{i,-L}^{K} + \theta\,\text{SIC2Ret}_{i,-L} + \psi\,\text{MktRet}_{-L} + \Lambda'\mathbf{X}_i + \mu_{j(i)} + \tau_{t(i)} + \varepsilon_i, \tag{1}$$

where K is the number of connected firms and L the return window in months. The vector $\mathbf{X}_i$ contains the log of the midpoint of the initial filing range, $\log(1 + \text{age})$, indicators for venture-capital or growth-capital backing and Internet firms, and exchange fixed effects. The specification also includes industry and offer-year fixed effects. Using the initial range midpoint avoids conditioning on the final offer price, a pricing outcome that connected returns may help explain. The coefficient β measures the information in connected returns beyond contemporaneous industry and market movements and the included issuer characteristics.

The sample contains 3,954 IPOs with an observed initial filing range and complete controls. The baseline regressions winsorize continuous variables at the 1st and 99th percentiles; the timing analysis in Table 6 uses untrimmed return signals and log-price outcomes. Errors are clustered by SIC2 industry and offer month. Tables 7 and 9 report null-imposed wild cluster bootstrap tests for the bookbuilding-return coefficients and the total-repricing uncertainty interaction. Appendix F provides implementation details, multiple-testing adjustments, and sensitivity checks; Appendix B reports additional specifications.

Higher connected returns predict higher first-day IPO returns (Table 4). In the six-month specification, a one-standard-deviation increase in connected returns, 35.70 percentage points, is associated with a 10.06-percentage-point higher first-day return, compared with a sample mean of 25.2%. One-standard-deviation increases over the one-, three-, and twelve-month windows correspond to first-day returns higher by 6.89, 9.40, and 7.76 percentage points, respectively. Each coefficient is positive and significant at 5%. These magnitudes use the SD of the winsorized connected return in the corresponding estimation sample and do not condition on offer price revisions.

The results are similar with textual-similarity weights and with industry-by-offer-year fixed effects, which compare issuers within the same industry and issuance year. The positive first-day relation also holds across the connected set sizes and return windows reported in Appendix B. I next examine how connected returns relate to each stage of price adjustment.

Table 4: Connected Returns and First-Day IPO Returns

	(1) 1 month	(2) 3 months	(3) 6 months	(4) 12 months
Connected return	0.646**	0.454**	0.282**	0.117**
	(0.197)	(0.134)	(0.095)	(0.035)
SIC2 industry return	0.831***	0.308**	0.183***	0.147***
	(0.226)	(0.123)	(0.044)	(0.031)
Market return	-0.762**	-0.547*	-0.610**	-0.229
	(0.326)	(0.279)	(0.212)	(0.172)
Wild cluster bootstrap p-value	0.025	0.030	0.050	0.034
IPO controls	Yes	Yes	Yes	Yes
Exchange FE	Yes	Yes	Yes	Yes
SIC2 FE	Yes	Yes	Yes	Yes
Offer year FE	Yes	Yes	Yes	Yes
N	3,954	3,954	3,954	3,954
Adjusted R^2	0.276	0.284	0.275	0.258
Dependent variable mean	0.252	0.252	0.252	0.252

Notes: The dependent variable is the first CRSP closing price divided by the offer price minus one. Connected return is the equal-weighted cumulative return of the ten closest connected firms from the indicated number of calendar months before the recorded offer date through the market close immediately before the first public trading day. The SIC2 industry return and market return cover the same window. IPO controls are log of the midpoint of the initial filing range, log one plus age, VC or growth-capital backing, and an Internet indicator. All columns include SIC2, offer year, and exchange fixed effects. Continuous variables are winsorized at the 1st and 99th percentiles. Parentheses contain two-way CR2 standard errors clustered by SIC2 and offer month; stars use coefficient-specific Satterthwaite tests. The bootstrap p-values use products of industry and offer-month weights. Appendix F provides details. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

4.2 Offer Price Revisions and First-Day Returns

Investors receiving IPO allocations pay the offer price; the first-day close measures market value at the end of initial trading. A favorable revaluation of connected firms could be reflected in a higher offer price, a larger first-day return, or both. Decomposing the price change separates the estimated response at the offering from that during initial trading.

Let $Mid_i = (Low_i + High_i)/2$ denote the midpoint of the initial filing range. Writing $Offer_i$ for the offer price and $Close_{i,1}$ for the first-day closing price, I define

$$\begin{aligned} Revision_i &= \log \frac{Offer_i}{Mid_i}, \\ FirstDay_i &= \log \frac{Close_{i,1}}{Offer_i}, \\ TotalUpdate_i &= \log \frac{Close_{i,1}}{Mid_i} = Revision_i + FirstDay_i. \end{aligned} \tag{2}$$

The first-day outcome is the log counterpart of the simple return in Section 4.1. Log changes make the two stages additive. Table 5 translates the coefficients into percentage changes in

the corresponding price ratios for a one-standard-deviation increase in connected returns, using each window's winsorized regressor and estimation sample.

For each outcome, I estimate

$$\begin{aligned} Y_i = \alpha + \beta_Y \, \mathrm{ConnRet}^K_{i,-L} + \theta \, \mathrm{SIC2Ret}_{i,-L} + \psi \, \mathrm{MktRet}_{-L} \\ + \Lambda' \mathbf{X}_i + \mu_{j(i)} + \tau_{t(i)} + \varepsilon_i, \end{aligned} \tag{3}$$

where Y_i is $Revision_i$, $FirstDay_i$, or $TotalUpdate_i$. I estimate the specification separately for two-, three-, and six-month connected returns, with industry and market returns over the same windows. Each window begins relative to the offer date and ends at the close two trading sessions before first public trading, $T-2$. These regressions retain the same 3,954 IPOs and the controls and fixed effects of Section 4.1. The bookbuilding specification measures returns from the range filing through the $T-2$ close, uses 3,830 IPOs, and adds full bookbuilding duration without including pre-range returns. Within each window, all three outcome regressions use the same sample and regressors. I winsorize the revision and first-day components separately and sum them to obtain the total update. The estimated connected return coefficient for the total update therefore equals the sum of the coefficients for the two pricing stages.

Connected returns predict both offer price revisions and first-day returns across all four windows (Table 5). A one-standard-deviation increase in the six-month connected return is associated with a 3.79% higher offer price relative to the midpoint of the initial filing range and a 5.79% higher first-day close relative to the offer price. The corresponding total price-ratio effect is 9.81%. The six-month SD is 35.38 percentage points, slightly below the 35.70 points in Table 4 because this return ends at $T-2$. For a one-standard-deviation increase in bookbuilding returns, 11.51 percentage points, the estimated price-ratio effects are 1.51%, 4.24%, and 5.81%, respectively. Both pricing-stage coefficients are positive and significant at 5% in every window.

For the six-month window, the share of the estimated total log-price response reflected in the offer price revision is

$$\frac{\beta_{Revision}}{\beta_{TotalUpdate}} = 0.398. \tag{4}$$

Approximately two-fifths of the estimated six-month response is reflected in the offer price revision and the remainder in the first-day return. The corresponding revision shares are 26.5% for bookbuilding, 35.7% for two-month returns, and 38.8% for three-month returns.

Table 5: Pricing Effects of a One-Standard-Deviation Increase in Connected Returns

Price-ratio changes (%) for a one-SD increase Return window	SD (pp)	Revision	First-day	Total update	Revision share
Bookbuilding	11.51	1.51** (0.42) [3.65]	4.24** (1.51) [2.87]	5.81** (1.84) [3.24]	26.5%
2 months	14.98	2.75*** (0.53) [5.23]	5.00** (1.50) [3.42]	7.89*** (2.04) [4.01]	35.7%
3 months	20.52	3.64*** (0.55) [6.76]	5.80*** (1.51) [3.95]	9.64*** (2.02) [4.99]	38.8%
6 months	35.38	3.79** (1.05) [3.70]	5.79** (1.84) [3.23]	9.81** (2.93) [3.50]	39.8%

Notes: Each row reports separate IPO-level regressions of log offer-price revision, log first-day return, and total log price update on the connected return over the indicated window. The midpoint of the initial filing range is $Mid_i = (Low_i + High_i)/2$; outcomes are $\log(Offer_i/Mid_i)$, $\log(Close_{i,1}/Offer_i)$, and their sum. Connected returns equally weight eligible firms among the ten closest matches. All return windows end at the close two trading sessions before first public trading ($T-2$). Monthly windows begin the indicated number of calendar months before the offer date; bookbuilding begins on the range filing date. Monthly rows use 3,954 IPOs each and bookbuilding uses 3,830. Controls include matching-window SIC2 and market returns, log of the midpoint of the initial filing range, $\log(1 + \text{age})$, VC or growth-capital backing, and an Internet indicator, with exchange, SIC2, and offer-year fixed effects. Bookbuilding additionally controls for full bookbuilding duration and excludes pre-range returns. First-day regressions do not control for the revision. Existing 1st/99th-percentile winsorization is retained; the total update sums the separately winsorized revision and first-day components. Regressors and observations are identical across outcomes within each row, so log coefficients are additive. Let s_L denote the sample SD of the winsorized connected return in that row's estimation sample; the SD column reports $100s_L$ in percentage points. Entries are $100[\exp(\beta s_L) - 1]$; parentheses contain delta-method SEs, $100 s_L \exp(\beta s_L)\, SE(\beta)$, treating s_L as fixed. Brackets contain the regression t-statistics. Revision shares are ratios of log coefficients, $\widehat{\beta}_{Revision}/\widehat{\beta}_{TotalUpdate}$, and are invariant to the common SD scaling within a row. Standard errors cluster by SIC2 and offer month; stars use coefficient-specific Satterthwaite tests. The bookbuilding row excludes pre-range controls, which are included in Table 7. Appendix F describes the adjustment-share bootstrap. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$. The six-month revision-share 95% interval is [0.23, 0.61], from 2,000 two-way cluster-bootstrap draws.

These shares are calculated from log coefficients rather than the displayed percentage effects. They describe how the estimated relation with connected returns is divided between the price paid by investors receiving IPO allocations and their first-day return.

Hanley (1993) documents the positive relation between offer price revisions and initial returns. Relating both pricing stages to connected stock returns extends research on partial adjustment to public information (Bradley and Jordan, 2002; Lowry and Schwert, 2004; Edelen and Kadlec, 2005).[2]

The preceding results use cumulative returns over the full pre-offer window. To examine whether their timing matters, I divide the three- and six-month windows at their trading-

[2]Underwriting arrangements can also affect the two pricing stages. Busaba and Restrepo (2022) show that underwriting spreads influence adjustment from the midpoint of the initial filing range to the offer price and the aftermarket price.

session midpoints and include early and late connected returns jointly, together with market and industry returns over the corresponding halves, the existing IPO controls and fixed effects, and full bookbuilding duration. Table 6 reports the relations with offer-price revisions, first-day returns, and total price updates for the same 3,830 dated IPOs and $T-2$ endpoint used in the bookbuilding analysis. These regressions use untrimmed return signals and literal log-price outcomes. As in Table 5, the first-day regressions do not condition on the revision.

Returns in the later halves of both windows have stronger associations with first-day returns and total price updates than returns in the earlier halves. The late-minus-early first-day coefficient differences are 0.334 and 0.225 in the three- and six-month specifications, respectively ($p = 0.004$ and $p = 0.002$). Total-price responses are also significantly larger for later returns. This timing motivates isolating returns realized during bookbuilding in Section 4.3 to examine how public information enters the offer price and first-day trading.

4.3 Price Adjustment during Bookbuilding

The initial filing range provides a starting point for examining how public information is incorporated into IPO prices. I distinguish connected-firm returns realized before the range filing from those realized during bookbuilding, when underwriters can revise the initial pricing terms.

For the 3,830 IPOs with an observed range filing date and complete data, I construct two connected return measures. The pre-range return covers the six months preceding the range filing, excluding the filing date. The bookbuilding return begins on the range filing date and ends at the close two trading sessions before first public trading, $T-2$. This endpoint excludes the final trading session before listing, whose closing return may be realized after the offer price has been set. The connected portfolio is selected by the range filing and held constant during bookbuilding.

I estimate

$$\begin{aligned} \mathit{Revision}_i = \alpha &+ \beta_1 \,\mathrm{ConnRet}_i^{\mathit{PreRange}} + \beta_2 \,\mathrm{ConnRet}_i^{\mathit{Bookbuilding}} \\ &+ \psi_1 \,\mathrm{MktRet}_i^{\mathit{PreRange}} + \psi_2 \,\mathrm{MktRet}_i^{\mathit{Bookbuilding}} \\ &+ \theta_1 \,\mathrm{SIC2Ret}_i^{\mathit{PreRange}} + \theta_2 \,\mathrm{SIC2Ret}_i^{\mathit{Bookbuilding}} \\ &+ \kappa \,\log(1+\mathrm{Days}_i^{\mathit{Bookbuilding}}) + \Lambda'\mathbf{X}_i + \mu_{j(i)} + \tau_{t(i)} + \varepsilon_i, \end{aligned} \tag{5}$$

Market and industry returns are measured over the same two windows as connected returns.

Table 6: Timing of Connected Returns and IPO Price Adjustment

Outcome	Early	Late	Late − early	p-value
Panel A: 3-month window				
Offer-price revision	0.137	0.220	0.083	
	(0.048)	(0.051)	(0.090)	0.397
First-day return	0.104	0.438	0.334	
	(0.050)	(0.116)	(0.068)	0.004
Total update	0.241	0.658	0.417	
	(0.061)	(0.155)	(0.117)	0.015
Panel B: 6-month window				
Offer-price revision	0.026	0.177	0.151	
	(0.038)	(0.026)	(0.029)	0.004
First-day return	0.062	0.287	0.225	
	(0.038)	(0.067)	(0.037)	0.002
Total update	0.088	0.464	0.376	
	(0.074)	(0.089)	(0.050)	< 0.001

Notes: Each horizon uses the same 3,830 dated IPOs. Returns begin three or six calendar months before the recorded offer date and end at the $T-2$ close before first public trading. The first $\lfloor N/2 \rfloor$ market sessions form the early half, where N is the number of sessions in the window; the odd extra session is late. Both halves enter jointly. Outcomes are the literal log offer/initial-midpoint revision, log first-close/offer return, and log first-close/initial-midpoint total update. Return signals and outcomes are untrimmed. In contrast, Table 5 uses 1st/99th-percentile winsorization and sums the separately winsorized pricing components; its monthly-window regressions use 3,954 IPOs. The samples and transformations therefore differ across the two tables. Other controls retain their baseline transformations. Connected returns equally weight available original top-ten matches meeting 60% observed-session coverage in each half, without replacement or reranking. Controls are early/late market and SIC2 returns, log one plus full bookbuilding trading days, log initial midpoint, log one plus age, venture or growth-capital backing, and an Internet or high-technology indicator, with exchange, SIC2 and offer-year fixed effects. No separate pre-range return control or revision control enters. Panels A and B report coefficients on decimal returns, with two-sided p-values for late minus early. Parentheses contain two-way CR2 standard errors clustered by SIC2 and offer month; contrast p-values use Satterthwaite tests. Tests are unadjusted for multiple comparisons. Late-return revision shares, defined as the revision coefficient divided by the total-update coefficient, are 33.4% and 38.2% for three and six months. Their 95% Fieller intervals are [21.8%, 51.7%] and [31.2%, 53.4%], using the full joint covariance and Satterthwaite reference. Late-minus-early share differences have two-sided p-values of 0.188 and 0.693 under a joint delta approximation with a normal reference; this approximation can be unreliable near zero total responses. Results remain similar with separate pre-range return controls or baseline-winsorized pricing outcomes.

The industry portfolios exclude the issuer's thirty closest connected firms. The duration control measures the number of trading days from the range filing to first public trading; the remaining controls and fixed effects follow Section 4.1. I estimate the corresponding first-day return regression with and without the offer price revision. These specifications estimate the relation with connected returns conditional on contemporaneous market and industry movements and returns realized before the range filing.

Offer prices are revised upward when connected firms appreciate during bookbuilding

Table 7: Connected Returns During Bookbuilding, Offer Price Revisions, and First-Day Returns

	(1) Revision	(2) First-day	(3) First-day
Connected return, pre-range	0.066***	0.075***	0.041**
	(0.013)	(0.016)	(0.013)
Connected return, bookbuilding	0.133***	0.367***	0.299***
	(0.035)	(0.123)	(0.103)
Revision			0.514***
			(0.081)
Wild cluster bootstrap p, bookbuilding return	0.0094	0.0006	0.0040
Observations	3,830	3,830	3,830
Adjusted R^2	0.260	0.266	0.408

Notes: The unit of observation is an IPO with an observed range date, defined as the filing date of the initial filing range. The dependent variable is $Revision_i = \log(Offer_i/Mid_i)$ in column (1) and $FirstDay_i = \log(Close_{i,1}/Offer_i)$ in columns (2) and (3). The pre-range connected return is measured over the six months preceding the range date, excluding that date; the bookbuilding connected return runs from the range date through the close two trading sessions before the IPO's first public trading day ($T-2$). Connected, industry, and market returns use identical endpoints. The duration control retains the full range-to-trading period. Column (3) adds the offer price revision. All specifications include market returns and same-SIC2 returns over the same two windows (the same-SIC2 return averages the public firms in the issuer's two-digit SIC industry other than its thirty closest connected firms) and the log of one plus the number of trading days from the range filing to first public trading. IPO controls are the log of the midpoint of the initial filing range, log one plus issuer age, VC or growth-capital backing, and an Internet indicator; offer price and offer proceeds are excluded because the outcomes are functions of the offer price. All specifications include exchange, SIC2, and offer year fixed effects. Continuous variables are winsorized at the 1st and 99th percentiles on the full IPO panel. Parentheses contain two-way CR2 standard errors clustered by SIC2 and offer month. Bookbuilding-return p-values and stars use the null-imposed wild cluster bootstrap with SIC2 weights and both error-clustering margins in studentization. Other stars use Satterthwaite tests. Appendix F reports confidence intervals, multiple-testing adjustments, and sensitivity checks. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

(Table 7). A one-standard-deviation increase in connected bookbuilding returns is associated with an offer price approximately 1.53% higher relative to the initial range midpoint. The corresponding first-day closing price is approximately 4.23% higher relative to the offer price before conditioning on the revision and 3.44% higher after conditioning. The corresponding bootstrap p-values are 0.0094, 0.0006, and 0.0040. The conditional first-day relation remains positive when each industry or offer year is omitted in turn (Appendix F).

Connected returns predict first-day returns among issuers with the same offer price revision. This relation is consistent with incomplete offer-price adjustment to public information identified through business similarity.

Table 8 examines whether business descriptions add information beyond conventional ways of selecting comparables. The alternative sets are public firms in the issuer's four-digit

SIC industry; the ten firms in its two-digit SIC industry closest in market capitalization to its expected capitalization; and the ten firms closest across industries in pre-IPO size, profitability, and tangibility. Each comparison includes the alternative portfolio's pre-range and bookbuilding returns alongside the corresponding connected, industry, and market returns. All portfolios use the same return windows.

The conditional first-day relation is similar across these comparisons. Using each column's estimation-sample SD, a one-standard-deviation increase in connected bookbuilding returns is associated with first-day closing prices approximately 3.34%, 3.38%, and 3.56% higher relative to the offer price in the SIC4, size-matched, and characteristic-matched specifications, respectively. The connected return coefficient exceeds each alternative bookbuilding coefficient, with all three differences significant at 5%. The coefficient comparisons use common return units. Business descriptions identify information about first-day returns beyond these conventional comparables, with their incremental pricing information strongest at this stage.

Appendix Tables B.5 and B.6 compare text-based connections with public firms explicitly identified as competitors in pre-IPO disclosures. The two sets provide distinct pricing signals, and their relative information content depends in part on the comparable-firm eligibility universe. Section 6 examines their complementary information using post-listing return comovement.

The results are similar when bookbuilding returns end at $T-1$ or $T-3$ (Appendix Table B.4), including when the IPO sample is held fixed across endpoints. Connected returns continue to predict revisions and first-day returns, with the latter relation holding both before and after conditioning on the revision. The revision relation also remains significant at 5% when the connected portfolio contains between five and thirty firms (Appendix Table B.3).

Figure 1 plots the three relations after residualizing both variables on the other regressors. Each point is the mean adjusted outcome within one of twenty groups formed on the adjusted connected return; each line is the corresponding full-sample regression slope.

The results extend research on information in IPO prospectuses (Hanley and Hoberg, 2010) by showing that disclosure also identifies an informative external signal. Connected returns predict offer price revisions and remain informative about first-day returns conditional on the revision and conventional comparison portfolios.

Table 8: Connected Returns During Bookbuilding versus Industry and Alternative Comparison Sets Using the Same Return Windows

	(1) Baseline	(2) SIC4 industry	(3) Size-matched	(4) Characteristic-matched
Panel A: Revision				
Connected return, bookbuilding	0.133***	0.064	0.125**	0.118**
	(0.035)	(0.035)	(0.038)	(0.035)
Wild cluster bootstrap p-value	[0.0094]	[0.148]	[0.022]	[0.023]
Connected return, pre-range	0.066***	0.064***	0.065***	0.063**
	(0.013)	(0.008)	(0.013)	(0.017)
SIC2 industry return, bookbuilding	0.549***	0.454***	0.470***	0.486***
	(0.099)	(0.092)	(0.101)	(0.100)
SIC2 industry return, pre-range	0.011	-0.012	-0.007	0.003
	(0.021)	(0.032)	(0.019)	(0.024)
Alternative set return, bookbuilding		0.183**	0.086*	0.062*
		(0.068)	(0.039)	(0.028)
Alternative set return, pre-range		0.019	0.017	0.019
		(0.015)	(0.011)	(0.028)
Log bookbuilding days	-0.093***	-0.094***	-0.092***	-0.100***
	(0.009)	(0.010)	(0.009)	(0.008)
Difference: connected − alternative, bookbuilding		-0.118	0.038	0.055
		(0.072)	(0.071)	(0.043)
Wild cluster bootstrap p-value		[0.171]	[0.648]	[0.295]
Observations	3,830	3,324	3,812	3,426
Adjusted R^2	0.260	0.281	0.261	0.274
Panel B: First-day return, conditional on Revision				
Connected return, bookbuilding	0.299***	0.285**	0.293**	0.304**
	(0.103)	(0.101)	(0.096)	(0.101)
Wild cluster bootstrap p-value	[0.0040]	[0.047]	[0.034]	[0.045]
Connected return, pre-range	0.041**	0.031*	0.040**	0.040*
	(0.013)	(0.014)	(0.014)	(0.015)
SIC2 industry return, bookbuilding	-0.056	-0.106	-0.125	-0.014
	(0.169)	(0.190)	(0.257)	(0.200)
SIC2 industry return, pre-range	0.053**	0.020	0.027	0.058*
	(0.019)	(0.025)	(0.028)	(0.026)
Alternative set return, bookbuilding		0.087**	0.071	-0.058
		(0.035)	(0.111)	(0.041)
Alternative set return, pre-range		0.048**	0.026	-0.008
		(0.019)	(0.017)	(0.018)
Log bookbuilding days	-0.003	-0.002	-0.001	-0.001
	(0.007)	(0.008)	(0.007)	(0.007)
Revision	0.514***	0.520***	0.514***	0.529***
	(0.081)	(0.085)	(0.080)	(0.086)
Difference: connected − alternative, bookbuilding		0.198**	0.222**	0.362**
		(0.076)	(0.071)	(0.100)
Wild cluster bootstrap p-value		[0.042]	[0.029]	[0.023]
Observations	3,830	3,324	3,812	3,426
Adjusted R^2	0.408	0.412	0.408	0.414

Notes: Panel A reports log offer-price revisions; Panel B reports log first-day returns conditional on the revision. Return regressors are in decimal units. Pre-range returns cover the six months before the range filing; bookbuilding returns run from that filing through $T-2$. Column (1) repeats the baseline in Table 7. Columns (2)–(4) add SIC4 industry returns, returns on ten size-matched firms within the issuer's SIC2 industry, and returns on ten firms matched across industries by assets, profitability, and tangibility, respectively. The difference row is the connected-minus-alternative bookbuilding coefficient. All columns retain the market and SIC2 returns, offering controls, duration control, and fixed effects of Table 7. Continuous variables are winsorized at the 1st and 99th percentiles. Parentheses contain two-way CR2 standard errors clustered by SIC2 and offer month. The connected bookbuilding-return tests in column (1) use the null-imposed bootstrap in Table 7. Elsewhere, stars use Satterthwaite tests and bootstrap p-values use products of industry and offer-month weights. Appendices E and F provide construction and inference details. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

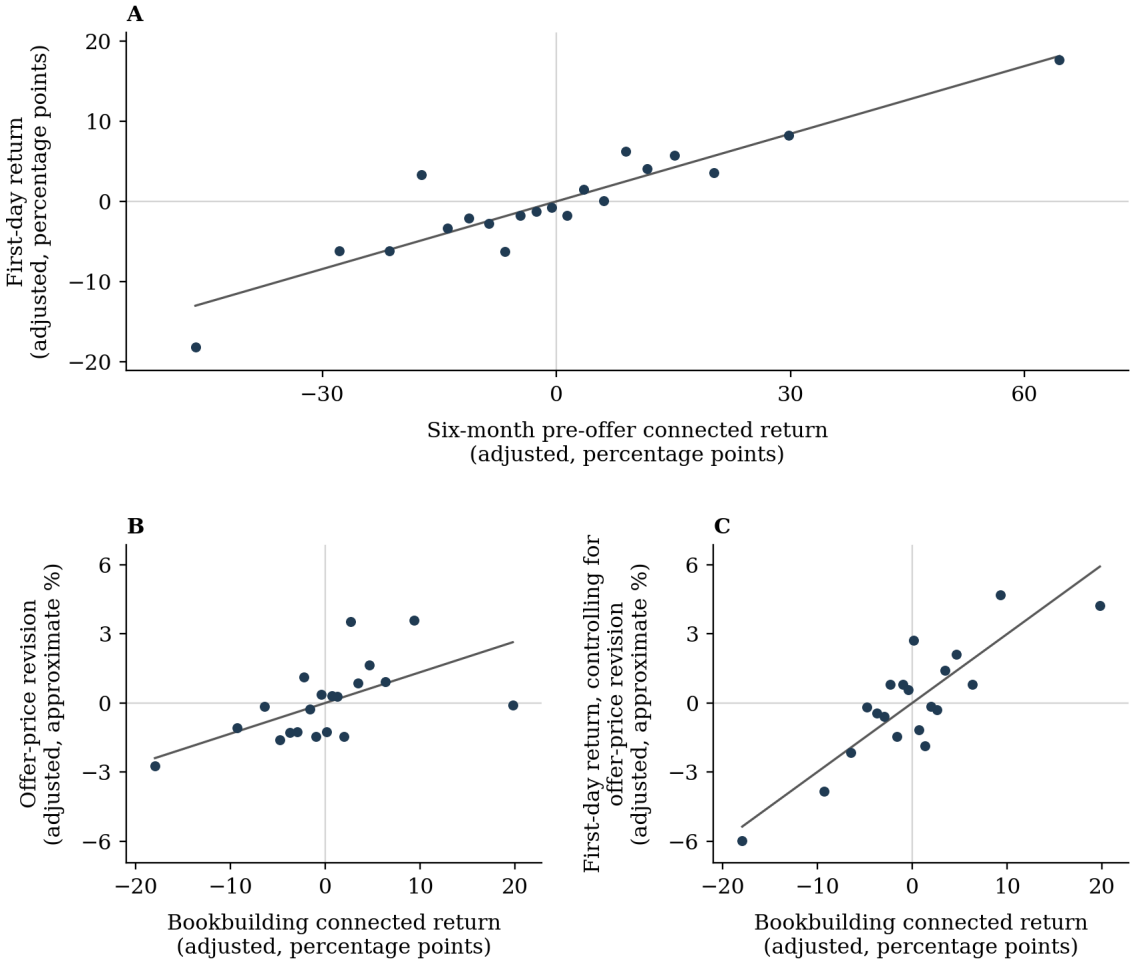


Figure 1: Connected Stock Returns and IPO Price Formation. Each panel plots mean adjusted connected returns and IPO outcomes within 20 approximately equal-sized groups formed separately on the adjusted connected return. Both variables are residualized on the other regressors, including the intercept and fixed effects, in the corresponding specification. Lines show the full-sample partial regression slopes. Panel A relates the simple first-day return to the six-month pre-offer connected return, following the six-month column of Table 4 ($N = 3{,}954$). Panel B relates the log offer price revision to connected returns during bookbuilding, controlling for pre-range connected returns and the other regressors in column (1) of Table 7. Panel C relates the log first-day return to connected returns during bookbuilding, conditional on the revision, as in column (3) of Table 7. Panels B and C use the same 3,830 IPOs and the same regressors, with Panel C adding the revision. Bookbuilding returns run from the range filing through the close two trading sessions before first public trading ($T - 2$). All panels use $K = 10$ and the winsorization of the corresponding regression. Adjusted variables are multiplied by 100; simple returns are in percentage points, while scaled log price changes approximate percentage changes in the corresponding price ratios. Slopes and 95% confidence intervals are 0.282 [0.040, 0.523], 0.133 [0.043, 0.229], and 0.299 [0.077, 0.481] in Panels A–C, respectively. Panel A uses SIC2/offer-month CR2/Satterthwaite inference; Panels B and C use the null-imposed wild cluster bootstrap in Table 7.

4.4 Valuation Uncertainty and Connected Announcement News

When information about the issuer's own prospects is less precise, news about economically related public firms may play a larger role in updating its valuation. Earnings announcements provide identifiable public-information events through which to examine this possibility. If filing uncertainty captures differences in the precision of investors' prior valuations, the same connected-firm announcement signal may be associated with a larger total price response for issuers with more uncertain filings.

Following Loughran and McDonald (2013), I measure uncertainty using language in the initial S-1/F-1. The filing precedes the announcement-news window, and I standardize its uncertainty-word frequency in the common estimation sample.

I measure connected news using returns around earnings announcements by the issuer's connected public firms during bookbuilding. The connected set and portfolio weights are fixed at the initial range filing. Appendix B.5 describes the announcement-return signal, filing uncertainty, and sample construction. I estimate

$$Y_i = \alpha + \beta_Y S_i^{EA} + \delta_Y Z_i + \gamma_Y \left(S_i^{EA} \times Z_i\right) + \Gamma_Y' \mathbf{X}_i + \varepsilon_i, \tag{6}$$

where S_i^{EA} is connected announcement news and Z_i is standardized filing uncertainty. The dependent variable is the offer price revision, the first-day return, or their sum, using the log-price definitions in equation (2). The control vector includes market and industry returns, their interactions with uncertainty, and measures of announcement exposure and data coverage. All three outcome regressions use the same observations and regressors.

Table 9 reports estimates for 3,632 IPOs. For a one-standard-deviation increase in connected announcement news, the estimated percentage change in price from the initial range midpoint to the first-day close is approximately 2.27 percentage points larger when filing uncertainty is one standard deviation higher. The total-repricing interaction coefficient is 1.165 (bootstrap $p = 0.0040$; 95% interval [0.551, 2.279]). The interaction is also positive for offer price revisions, at 0.696 ($p = 0.059$), and first-day returns, at 0.469 ($p = 0.074$), although these individual-stage estimates are less precise. Appendix F reports the industry- and year-omission checks.

These results are consistent with connected public information being more useful when the issuer is harder to value. Uncertain language may also reflect differences in economic exposure or disclosure choices, so the estimates do not isolate a causal effect of valuation uncertainty.

Table 9: Filing Uncertainty and the Response to Connected Announcement News

	(1) Offer-price revision	(2) First-day return	(3) Total repricing
Connected announcement news	0.409	0.156	0.566
	(0.428)	(0.205)	(0.514)
Filing uncertainty	-0.012	0.005	-0.006
	(0.007)	(0.013)	(0.017)
Connected announcement news × uncertainty	0.696*	0.469*	1.165***
	(0.299)	(0.218)	(0.362)
Interaction p-value	0.059	0.074	0.0040
Additional response to one-SD news for one-SD higher uncertainty (pp, approx.)	1.35	0.91	2.27
Market and industry return interactions with uncertainty	Yes	Yes	Yes
Announcement exposure and coverage controls	Yes	Yes	Yes
Observations	3,632	3,632	3,632
Adjusted R^2	0.270	0.255	0.315

Notes: The outcomes are log offer price relative to the initial range midpoint, log first-day close relative to offer price, and their sum. Connected announcement news aggregates adjusted connected-firm returns in eligible three-session earnings windows during bookbuilding. Filing uncertainty is measured from the initial S-1/F-1 and standardized. All columns use identical observations and regressors, including both main effects, market and industry returns and their interactions with uncertainty, nonannouncement returns and their uncertainty interaction, exposure and coverage controls, and the remaining controls and fixed effects in Appendix B.5. The announcement-news main effects are evaluated at the common-sample mean industry and similarity composition.

The economic-effect row is $100\widehat{\gamma}_Y s_{EA}$, using the same common-sample announcement-news SD, $s_{EA} = 0.019478$, in all columns. It approximates the difference in the percentage price response to one-SD news associated with one-SD higher uncertainty, in percentage points. Parentheses contain two-way CR2 standard errors clustered by SIC2 and offer month. For the total-repricing interaction in column (3), the p-value and stars use the null-imposed wild cluster bootstrap with SIC2 weights and both error-clustering margins in studentization. Other p-values and stars use Satterthwaite tests. Appendix F reports confidence intervals, multiple-testing adjustments, and sensitivity checks. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

5 Connected Firms and Issuer Financing Around the IPO

A firm's capital structure reflects both its underlying financing needs and the financing choices available to it over time. Public firms with similar businesses may provide a benchmark for those needs because products, assets, and growth opportunities can affect borrowing capacity and financial distress costs. An issuer's pre-IPO leverage, however, also reflects its financing history and access to private capital. Going public creates an opportunity to revise this inherited position. Issuers with leverage above that of connected firms may use their new access to equity to reduce outstanding debt, whereas issuers below the benchmark need

not increase borrowing immediately. This motivates examining debt adjustments separately for issuers starting above and below connected leverage, conditional on initial leverage and conventional financing determinants.

I first study leverage at listing and subsequent leverage conditional on the issuer's pre-IPO position, then compare debt changes for issuers starting above and below connected leverage.

5.1 Leverage at Listing

I measure leverage at listing at the first fiscal quarter-end on or after the offer date, within 270 days. Book leverage is current plus long-term debt divided by book assets; market leverage is debt divided by the sum of debt and market equity, measured at the month-end closest to the accounting date. Financial firms and utilities are excluded. Connected leverage is the equal-weighted average leverage of the ten public firms selected using business descriptions disclosed before the range filing, as described in Section 3. Their accounting measures use the latest eligible annual statements preceding the offering. Appendix E provides the accounting dates and construction details.

The baseline specification is

$$\begin{aligned} \text{Leverage}_{ijt}^{IPO} = \alpha + \beta\, \text{ConnLev}_{i,t-1}^{K=10} + \theta\, \text{IndLev}_{j,t-1} + \Gamma' \overline{\mathbf{Z}}_{i,t-1}^{K=10} \\ + \Lambda' \mathbf{X}_{it} + \Psi' \mathbf{W}_{i,t-1} + \mu_j + \tau_t + \varepsilon_{ijt}, \end{aligned} \tag{7}$$

where i indexes issuers, j two-digit SIC industries, and t offer years. The industry benchmark is the equal-weighted average leverage of public firms in the issuer's SIC2 industry. The controls include connected firms' size, market-to-book, profitability, tangibility, cash, and R&D intensity ($\overline{\mathbf{Z}}$); offering characteristics ($\mathbf{X}$); and the issuer's pre-IPO size, profitability, and tangibility ($\mathbf{W}$). Industry, offer-year, and exchange fixed effects account for broader differences across issuers. The coefficient β measures the information in connected leverage beyond these conventional financing determinants.[3]

Financing regressions cluster errors by SIC2 industry and offer year. Tables 12 and 13 report null-imposed wild cluster bootstrap tests for the conditional one-year leverage coefficient and the listing-quarter connected-gap coefficients and contrasts. Appendix F reports multiple-testing adjustments and alternative inference procedures.

[3]Offering controls are log reported gross proceeds, $\log(1 + \text{age})$, Internet and dual-class indicators, and offer price. Continuous variables are winsorized at the 1st and 99th percentiles. Before computing connected and industry averages, firm-level leverage is winsorized within fiscal year across the Compustat universe, and book leverage is capped at one. Appendix E details the construction of the connected and industry averages.

Table 10 shows that issuers connected to more highly levered public firms have higher leverage at listing. In column (4), one-standard-deviation increases in connected book and market leverage are associated with issuer book and market leverage higher by approximately 1.52 and 1.95 percentage points, respectively. The connected market leverage SD is 11.21 percentage points. Throughout the paper, I scale coefficients in unstandardized leverage units using the common SDs in Table 1. These differences compare with sample means of 15.1% and 9.3%, respectively. The book estimate is significant at 10% and the market estimate at 5%. The coefficients remain positive when industry-by-offer-year fixed effects replace separate industry and year effects in column (6), indicating that the relation also holds among issuers entering public markets in the same industry and year.

Table 11 compares the connected benchmark with industry leverage and accounting-based matches. The alternative comparables are public firms in the issuer's two-digit SIC industry and the five firms closest across industries based only on pre-IPO size, profitability, and tangibility. Connected market leverage remains informative alongside either benchmark: a one-standard-deviation increase predicts approximately 1.9 percentage points higher issuer market leverage in each comparison. Both estimates are significant at 5%. A one-standard-deviation increase in characteristic-matched leverage predicts book and market leverage higher by 2.5 and 1.2 percentage points, respectively. Business descriptions capture financing similarities beyond industry membership and proximity in these accounting characteristics.

5.2 Financing Information Beyond Initial Leverage

I examine whether connected leverage predicts post-IPO capital structure beyond the issuer's pre-IPO leverage.

Table 12 follows issuer book leverage from the last pre-IPO quarter through listing and the first two years afterward. Connected leverage is measured before the offering and held fixed across horizons. Panel A uses the controls in column (4) of Table 10. Panel B adds pre-IPO leverage to compare subsequent capital structures among issuers with similar starting leverage.

Financing similarities are evident before the offering: the pre-IPO connected-leverage coefficient in Panel A is 0.365. Conditional on the issuer's pre-IPO leverage, a one-standard-deviation increase in connected book leverage is associated with issuer book leverage higher

Table 10: Connected Firm Leverage and IPO Leverage at Listing

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: IPO book leverage at listing						
Connected firm book leverage	0.697***	0.473***	0.229***	0.129*	0.175**	0.165**
	(0.077)	(0.048)	(0.039)	(0.055)	(0.057)	(0.051)
Wild cluster bootstrap p-value	[0.004]	[0.003]	[0.010]	[0.058]	[0.026]	[0.038]
SIC2 industry benchmark		0.173	0.175	0.193	0.446***	0.907
		(0.241)	(0.181)	(0.191)	(0.057)	(1.854)
IPO controls	No	No	Yes	Yes	Yes	Yes
Issuer pre-IPO characteristics	No	No	Yes	Yes	Yes	Yes
Connected firm controls	No	No	No	Yes	Yes	Yes
SIC2 FE	No	Yes	Yes	Yes	No	No
Offer year FE	Yes	Yes	Yes	Yes	Yes	No
Exchange FE	No	No	Yes	Yes	Yes	Yes
SIC2 × offer year FE	No	No	No	No	No	Yes
N	3,728	3,728	3,728	3,728	3,728	3,728
Adjusted R^2	0.160	0.223	0.387	0.392	0.383	0.452
Dependent variable mean	0.151	0.151	0.151	0.151	0.151	0.151
Panel B: IPO market leverage at listing						
Connected firm market leverage	0.707***	0.535***	0.227***	0.174**	0.228***	0.230**
	(0.063)	(0.060)	(0.037)	(0.067)	(0.067)	(0.069)
Wild cluster bootstrap p-value	[0.002]	[0.003]	[0.009]	[0.028]	[0.009]	[0.028]
SIC2 industry benchmark		-0.003	-0.032	-0.027	0.216***	-0.159
		(0.069)	(0.073)	(0.079)	(0.028)	(0.534)
IPO controls	No	No	Yes	Yes	Yes	Yes
Issuer pre-IPO characteristics	No	No	Yes	Yes	Yes	Yes
Connected firm controls	No	No	No	Yes	Yes	Yes
SIC2 FE	No	Yes	Yes	Yes	No	No
Offer year FE	Yes	Yes	Yes	Yes	Yes	No
Exchange FE	No	No	Yes	Yes	Yes	Yes
SIC2 × offer year FE	No	No	No	No	No	Yes
N	3,726	3,726	3,726	3,726	3,726	3,726
Adjusted R^2	0.269	0.326	0.553	0.555	0.536	0.634
Dependent variable mean	0.093	0.093	0.093	0.093	0.093	0.093

Notes: The outcomes are issuer book and market leverage at listing: debt divided by assets and by debt plus market equity, respectively. Connected firm leverage is the equal-weighted average for the ten closest firms by business description similarity. The SIC2 industry benchmark averages public firm leverage in the issuer's two-digit SIC industry in the prior fiscal year, excluding financial firms and utilities. IPO controls are log reported gross proceeds, log(1 + age), Internet and dual-class indicators, and offer price. Issuer pre-IPO characteristics are profitability, log assets, and tangibility; connected firm controls are average size, market-to-book, profitability, tangibility, cash, and R&D intensity. Column (6) replaces separate SIC2 and offer year fixed effects with SIC2-by-offer-year fixed effects. The industry benchmark can vary within these cells because its accounting cutoff depends on the offer date. Book and market leverage specifications use fixed samples of 3,728 and 3,726 IPOs, respectively; singleton cells are retained. Section 5 and Appendix E describe measurement timing and winsorization. Parentheses contain two-way CR2 standard errors clustered by SIC2 and offer year; stars use Satterthwaite tests. Brackets report bootstrap p-values using products of industry and offer-year weights. Appendix F provides details. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table 11: Connected Firm Leverage versus Alternative Comparable Sets

	(1) Book	(2) Market	(3) Book	(4) Market
Connected firm leverage	0.015*	0.019**	0.014*	0.019**
	(0.006)	(0.007)	(0.006)	(0.008)
Wild cluster bootstrap p-value	[0.058]	[0.028]	[0.062]	[0.034]
Industry leverage	0.013	-0.002		
	(0.013)	(0.007)		
Wild cluster bootstrap p-value	[0.276]	[0.729]		
Characteristic-matched			0.025***	0.012***
			(0.003)	(0.003)
Wild cluster bootstrap p-value			[0.004]	[0.005]
Difference: connected − alternative	0.002	0.022*	-0.011	0.007
	(0.012)	(0.012)	(0.007)	(0.009)
Wild cluster bootstrap p-value	[0.863]	[0.072]	[0.181]	[0.406]
p-value: connected = alternative	0.887	0.097	0.192	0.468
Additional industry control	No	No	Yes	Yes
IPO controls	Yes	Yes	Yes	Yes
Issuer pre-IPO characteristics	Yes	Yes	Yes	Yes
Connected firm controls	Yes	Yes	Yes	Yes
Industry, offer year, and exchange FE	Yes	Yes	Yes	Yes
N	3,728	3,726	3,728	3,726
Adjusted R^2	0.392	0.555	0.401	0.558
Dependent variable mean	0.151	0.093	0.151	0.093

Notes: The outcomes are issuer book and market leverage at listing. Connected firm leverage is the equal-weighted average for the ten closest firms by business description similarity. Both alternatives draw on the same public firm universe in the prior fiscal year with statements dated on or before the offer, excluding financial firms and utilities. Industry leverage (columns (1) and (2)) is the equal-weighted average for firms with the issuer's two-digit SIC code, using the industry measure in Table 10. Characteristic-matched (columns (3) and (4)) is the equal-weighted leverage of the 5 firms closest to the issuer across industries, matched only on pre-IPO size (log assets), profitability, and tangibility (Euclidean distance; each variable standardized by its fiscal-year mean and standard deviation among public firms); the issuer is excluded from the matched set. Connected and alternative leverage are standardized on each column's estimation sample, so coefficients are the change in issuer leverage per one-standard-deviation increase in the benchmark. The difference row reports $\beta_T - \beta_M$ from the reparameterized regression, with its own standard error and p-values. Each column uses the complete cases for its own alternative within the fixed book- and market-leverage samples. Controls and fixed effects follow column (4) of Table 10; industry leverage enters once as the standardized alternative in columns (1) and (2), and as an additional unstandardized control in columns (3) and (4). Parentheses contain two-way CR2 standard errors clustered by SIC2 and offer year; stars use Satterthwaite tests. Brackets report bootstrap p-values using products of industry and offer-year weights. Appendix F provides details. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

by 2.04 percentage points after one year and 2.55 percentage points after two years. The one-year coefficient is 0.173 (bootstrap $p = 0.0359$; 95% interval [0.013, 0.315]); the two-year estimate is significant at 10%. The listing-quarter coefficient is 0.064, with a standard error of 0.047. Appendix Table A.2 reports the comparison requiring the same issuers at all four dates, and Appendix F reports alternative inference procedures.

The post-IPO relation therefore extends beyond persistence in issuers' own leverage.

5.3 Connected Leverage and Debt Adjustments Around and After the IPO

The preceding results concern the capital structure issuers maintain after listing. I next examine whether connected leverage helps explain changes in outstanding debt around the IPO. The relation need not be symmetric. Issuers with relatively high leverage may use access to public equity to reduce debt and restore borrowing capacity, whereas issuers starting below the connected benchmark may preserve financial slack for future investment. These different financing incentives motivate examining debt adjustments separately above and below connected leverage (Leary and Roberts, 2005; Faulkender et al., 2012).

Among issuers whose pre-IPO leverage exceeds connected leverage, the gap often predates the offering in the subsample with historical accounting data. Of those with an accounting statement one year earlier, 85.2% also exceeded the connected benchmark at that time; the corresponding figure two years earlier is 78.3% (Appendix Table A.8). Both comparisons hold connected leverage fixed at its IPO value.

Leverage can decline as assets grow even when debt does not fall. Alongside changes in book leverage, I therefore examine changes in debt divided by fixed pre-IPO assets. This measure isolates changes in the numerator of the leverage ratio. Debt outstanding falls for approximately two-thirds of issuers in the listing sample.

I define nonnegative distances above and below connected leverage as

$$\text{AboveGap}_i = \max(\text{PreLev}_i - \text{ConnLev}_i, 0), \qquad \text{BelowGap}_i = \max(\text{ConnLev}_i - \text{PreLev}_i, 0).$$

Both leverage measures are book leverage, and connected leverage is measured before the offering. Each distance is positive only for issuers on the corresponding side of the benchmark.

Table 12: Connected Firm Leverage and Issuer Leverage Before and After the IPO

Issuer book leverage measured at:	Pre-IPO (1)	Listing (2)	+1 year (3)	+2 years (4)
Panel A: Without issuer pre-IPO leverage				
Connected firm book leverage	0.365***	0.125*	0.229**	0.263**
	(0.097)	(0.056)	(0.071)	(0.100)
Wild cluster bootstrap p-value	[0.016]	[0.066]	[0.031]	[0.048]
Observations	3,699	3,699	3,514	3,160
Adjusted R^2	0.167	0.395	0.382	0.297
Panel B: Conditional on issuer pre-IPO leverage				
Connected firm book leverage		0.064	0.173**	0.217*
		(0.047)	(0.069)	(0.100)
Wild cluster bootstrap p-value		[0.209]	[0.0359]	[0.089]
Issuer pre-IPO book leverage		0.168***	0.154***	0.154***
		(0.022)	(0.014)	(0.014)
Observations		3,699	3,514	3,160
Adjusted R^2		0.516	0.481	0.372
Dependent variable mean	0.385	0.150	0.174	0.206
Dependent variable standard deviation	0.496	0.219	0.224	0.254
Median days from offer to statement date	-42	49	414	779

Notes: The outcome in every column is the issuer's book leverage, current plus long-term debt divided by book assets, measured at the last fiscal quarter-end strictly before the offer (column 1), the first quarter-end on or after the offer within 270 days (column 2), and the quarter-ends closest to the listing statement date plus one and two years, within 60 days (columns 3 and 4). The regressor of interest is the same pre-IPO connected firm book leverage in every column: the equal-weighted average for the ten closest firms by business description similarity, with accounting periods ending no later than the offer date. All columns use the specification of column (4) of Table 10: the same-SIC2 benchmark, IPO controls, issuer pre-IPO profitability, size, and tangibility, connected firm controls, and SIC2, offer year, and exchange fixed effects. Panel B adds the issuer's pre-IPO book leverage, so the connected-leverage coefficient compares post-IPO leverage across issuers with the same starting leverage and other included characteristics. Equivalently, it is the coefficient on connected leverage in a regression of the change from pre-IPO leverage on the same regressors. Each column uses the issuers for which its outcome, the issuer's pre-IPO leverage, and all regressors are observed, so Panels A and B compare identical samples within a column; Appendix Table A.2 holds the sample fixed across columns. Outcomes are winsorized at the 1st and 99th percentiles. Parentheses contain two-way CR2 standard errors clustered by SIC2 and offer year. For connected leverage in Panel B, column (3), the bracketed p-value and stars use the null-imposed wild cluster bootstrap with offer-year weights and both error-clustering margins in studentization. Elsewhere, stars use Satterthwaite tests and brackets report bootstrap p-values using products of industry and offer-year weights. Appendix F reports confidence intervals, multiple-testing adjustments, and sensitivity checks. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

The starting specification is

$$\begin{aligned} Y_{ih} = \alpha_h + \beta_h^A \text{AboveGap}_i + \beta_h^B \text{BelowGap}_i + \gamma_h \text{PreLev}_i \\ + \boldsymbol{\delta}_h' \mathbf{C}_i + \mu_{j,h} + \tau_{t,h} + \eta_{e,h} + \varepsilon_{ih}, \end{aligned} \tag{8}$$

where Y_{ih} is the change in book leverage or the change in debt divided by pre-IPO assets, and $\mathbf{C}_i$ contains industry leverage and the issuer, connected-firm, and offering controls from Section 5.1. The remaining terms are industry, offer-year, and exchange fixed effects. Outcomes are measured from the pre-IPO quarter to listing or to the subsequent one- and two-year horizons, with accounting dates defined in Section 5.2. Conditional on starting leverage, the distance coefficients compare issuers with different connected benchmarks, separately above and below the benchmark. A negative β_h^A indicates greater debt or leverage reductions for issuers starting farther above connected leverage. A common slope in connected leverage imposes $\beta_h^A + \beta_h^B = 0$, because a higher benchmark reduces the distance above it and increases the distance below it.

The specifications in Table 13 model pre-IPO leverage with a four-knot restricted cubic spline. Panel A is the baseline: it uses observed connected leverage and includes nonnegative distances above and below industry leverage. Panel B is a conditional robustness check that adds corresponding distances relative to predicted leverage from a public-firm model using size, profitability, tangibility, and industry leverage, evaluated at the issuer's pre-IPO characteristics (Appendix A.7). Benchmark levels are omitted where they are redundant with the distances and initial leverage. These controls allow financing changes to vary nonlinearly with initial leverage and account for differences captured by conventional financing benchmarks. The panels use identical samples within each horizon, and the spline knots remain fixed.

I scale economic magnitudes using the Table 1 SD of connected book leverage, holding issuer pre-IPO leverage and the other controls fixed. Each comparison stays within the same pre-IPO above- or below-benchmark region. The scaling uses variation in connected leverage rather than the SD of either nonnegative gap.

Among issuers starting above the connected benchmark, a one-standard-deviation lower benchmark is associated with additional debt reductions equivalent to 2.98% of pre-IPO assets in the observed-benchmark baseline. The baseline coefficient is -0.252 (bootstrap $p = 0.0102$; 95% interval $[-0.420, -0.070]$). The economic effect is 2.73% after adding predicted-leverage gaps on the same sample, with the fitted benchmark treated as fixed. The negative relation persists when each industry or offer year is omitted in turn (Appendix F).

The corresponding below-benchmark debt-change estimate is small relative to the above-benchmark estimate. The bootstrap rejects the common-slope restriction, $\beta_h^A + \beta_h^B = 0$, with $p = 0.0109$. The estimates are consistent with issuers using the offering to reduce relatively high debt while less levered issuers preserve financial slack. Appendix F reports multiple-testing adjustments and sensitivity to alternative bootstrap weights.

The leverage-change results extend beyond the offering. Among issuers starting above the connected benchmark, a one-standard-deviation lower benchmark is associated with an additional decline in book leverage of 2.39 percentage points after one year and 2.89 percentage points after two years. Both estimates and the common-slope contrasts are significant at 10% under the product-weight wild bootstrap. Predicted leverage also helps explain subsequent leverage changes, indicating that connected leverage and conventional financing benchmarks both contribute information about issuer leverage after listing.

Appendix A.9 examines alternative measures of debt reduction relative to initial debt balances.

Table 14 examines the timing of debt changes. I compare the quarter preceding the offering (-2 to -1), the IPO quarter (-1 to 0), and the first four post-listing quarters (0 to $+4$) using the same 2,303 issuers with the required accounting statements. All debt changes are divided by the same pre-IPO asset balance. The specifications retain the controls of Table 13, Panel B, and add fixed effects for five-percentage-point intervals of starting leverage. For issuers initially above connected leverage, the debt-change relation is more negative during the IPO quarter than during the preceding quarter. For a one-standard-deviation lower connected benchmark within this region, the directly estimated difference between IPO-quarter and preceding-quarter debt changes is -5.09% of pre-IPO assets (wild cluster bootstrap $p = 0.055$).

During the first post-listing year, larger initial distances above the benchmark predict lower debt changes. For issuers initially below the benchmark, a one-standard-deviation higher connected benchmark predicts a post-listing debt change higher by 9.64% of pre-IPO assets (95% interval: -2.26% to 21.55%; wild bootstrap $p = 0.083$), compared with 0.39% around the offering. This timing is consistent with issuers reassessing financing needs and borrowing capacity after listing.

First-year debt issuance provides a complementary measure of financing activity for issuers

Table 13: Financing Changes Beyond Industry and Predicted Leverage

	Debt change/A_{pre}	Change in book leverage		
	Listing	Listing	+1 year	+2 years
	(1)	(2)	(3)	(4)
Panel A: Connected and industry gaps				
Connected leverage distances				
Below connected leverage (BelowGap_C)	-0.030	-0.086*	0.016	0.065
	(0.072)	(0.040)	(0.059)	(0.074)
Wild p-value (BelowGap_C)	0.681	0.064	0.804	0.452
Above connected leverage (AboveGap_C)	-0.252**	-0.097*	-0.226**	-0.273**
	(0.082)	(0.045)	(0.068)	(0.099)
Wild p-value (AboveGap_C)	0.0102	0.088	0.029	0.048
Connected slope sum ($\beta_C^A + \beta_C^B$)	-0.282	-0.183	-0.209	-0.208
95% confidence interval	[-0.502, -0.077]	[-0.324, -0.042]	[-0.373, -0.046]	[-0.351, -0.064]
Wild p-value ($\beta_C^A + \beta_C^B = 0$)	0.0109	0.030	0.026	0.030
N	3,697	3,697	3,512	3,158
R^2	0.652	0.902	0.887	0.826
Industry leverage gaps (issuer below/above)	Yes	Yes	Yes	Yes
Predicted leverage gaps (issuer below/above)	No	No	No	No
Flexible pre-IPO leverage	Yes	Yes	Yes	Yes
Panel B: Adding predicted leverage gaps				
Connected leverage distances				
Below connected leverage (BelowGap_C)	-0.000	-0.069	0.008	0.054
	(0.070)	(0.043)	(0.063)	(0.080)
Wild p-value (BelowGap_C)	0.998	0.118	0.909	0.544
Above connected leverage (AboveGap_C)	-0.231**	-0.081	-0.202**	-0.245*
	(0.076)	(0.049)	(0.075)	(0.104)
Wild p-value (AboveGap_C)	0.0136	0.169	0.057	0.080
Predicted leverage distances				
Below predicted leverage (BelowGap_P)	-0.368	-0.156	0.361**	0.462**
	(0.225)	(0.116)	(0.103)	(0.171)
Wild p-value (BelowGap_P)	0.167	0.181	0.013	0.031
Above predicted leverage (AboveGap_P)	0.004	-0.087	-0.535**	-0.650*
	(0.186)	(0.127)	(0.184)	(0.292)
Wild p-value (AboveGap_P)	0.985	0.514	0.022	0.065
Connected slope sum ($\beta_C^A + \beta_C^B$)	-0.232	-0.150	-0.194	-0.191
95% confidence interval	[-0.427, -0.024]	[-0.314, 0.013]	[-0.378, -0.009]	[-0.355, -0.027]
Wild p-value ($\beta_C^A + \beta_C^B = 0$)	0.0346	0.074	0.055	0.053
N	3,697	3,697	3,512	3,158
R^2	0.653	0.903	0.888	0.827
Industry leverage gaps (issuer below/above)	Yes	Yes	Yes	Yes
Predicted leverage gaps (issuer below/above)	Yes	Yes	Yes	Yes
Flexible pre-IPO leverage	Yes	Yes	Yes	Yes

Notes: Column (1) reports debt changes divided by pre-IPO assets; columns (2)–(4) report changes in book leverage. For benchmark B, $\text{AboveGap} = \max(\text{PreLev} - B, 0)$ and $\text{BelowGap} = \max(B - \text{PreLev}, 0)$. Both panels include connected and industry gaps, flexible pre-IPO leverage, and the controls and fixed effects of Section 5.1. Panel A is the observed-benchmark baseline; Panel B adds predicted-leverage gaps, treating the fitted benchmark as fixed (Appendix A.7). The panels use the same issuers within each column. The common-slope restriction is $\beta_C^A + \beta_C^B = 0$, since a higher benchmark reduces the above gap and increases the below gap. Parentheses contain two-way CR2 standard errors clustered by SIC2 and offer year. In column (1) of both panels, above-gap p-values and stars and slope-sum p-values and intervals use the null-imposed wild cluster bootstrap with offer-year weights and both error-clustering margins in studentization. Elsewhere, stars and intervals use Satterthwaite tests and bootstrap p-values use products of industry and offer-year weights. Appendix F reports multiple-testing adjustments and sensitivity checks. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table 14: Connected Leverage and Debt Adjustments Around and After the IPO

	Above connected leverage			Below connected leverage		
Period	Effect	95% interval	Wild p	Effect	95% interval	Wild p
Preceding quarter (−2 to −1)	+1.30	[-0.10, 2.71]	0.074	+0.28	[-0.39, 0.95]	0.273
IPO quarter (−1 to 0)	-3.04	[-6.60, 0.52]	0.101	+0.39	[-1.14, 1.91]	0.515
Post-listing (0 to +4)	-6.50	[-10.83, -2.17]	0.018	+9.64	[-2.26, 21.55]	0.083
IPO minus preceding quarter	-5.09	[-9.91, -0.27]	0.055	+0.85	[-1.07, 2.76]	0.287

Notes: Effects are percentages of pre-IPO assets for a one-standard-deviation change in connected book leverage (11.80 percentage points, Table 1): a lower benchmark for issuers starting above it and a higher benchmark for those starting below it, holding initial leverage and other controls fixed. Negative entries indicate lower debt changes. All rows use the same 2,303 issuers. Quarter −1 is the last pre-offer quarter-end, quarter 0 the next scheduled quarter-end, and +4 four fiscal quarters later. The final row estimates the IPO-minus-preceding-quarter contrast directly. Controls follow Table 13, Panel B, with five-percentage-point initial-leverage fixed effects. Intervals use two-way CR2/Satterthwaite inference clustered by SIC2 and offer year; two-sided bootstrap p-values use products of industry and offer-year weights. Appendices E and F provide construction and inference details.

initially below connected leverage. In the sample with observed first-year debt cash flows, a ten-percentage-point larger initial gap below connected leverage is associated with additional gross long-term debt issuance equivalent to 5.29% of pre-IPO assets during the first year after the listing-quarter accounting date (wild cluster bootstrap $p = 0.015$), conditional on initial leverage, conventional leverage benchmarks, and firm characteristics. Appendix Table A.11 reports these estimates using winsorized cash-flow outcomes. The corresponding net long-term debt issuance estimate is positive but not significant at 5% (wild cluster bootstrap $p = 0.064$). The gross-issuance evidence complements the debt-balance results in Table 14 by linking connected leverage to subsequent borrowing activity. Among issuers initially above connected leverage, a larger initial gap is associated with lower gross and net long-term debt issuance during the following year (Appendix Table A.11).

Among issuers whose pre-IPO leverage exceeds connected leverage at the start of the information window, higher connected financing news is associated with lower long-term debt repayments during the first post-listing year (Appendix Table A.12). The repayment coefficient is -1.238 ($p = 0.014$). Gross borrowing also has a negative point estimate, while the net-borrowing estimate is positive and imprecise.

Pagano et al. (1998) find that Italian firms use IPOs to rebalance financing after earlier investment, while Alti (2006) documents continued capital structure adjustment after the initial equity issue. The leverage of related public businesses helps explain which issuers make larger financing adjustments, conditional on initial leverage and conventional financing characteristics.

Connected leverage thus predicts both post-listing capital structure and debt adjustment

around the offering. Among issuers starting above the benchmark, larger initial distances are associated with greater debt reductions relative to pre-IPO assets.

6 Business Similarity and Return Comovement

The usefulness of comparables selected from business descriptions depends on whether textual similarity identifies economically meaningful relationships. Firms operating in related markets may face common changes in demand, costs, technologies, and competitive conditions, causing their stock returns to move together even beyond industry-wide movements. Since issuer returns play no role in selecting connected firms, post-listing comovement provides an assessment of the economic relevance of the pre-IPO matches. However, return comovement alone does not establish shared operating exposures. I therefore complement the portfolio-return analysis with tests of factor- and industry-adjusted pairwise return correlations and operating-performance comovement.

6.1 Comovement Beyond Industry Returns

I measure monthly issuer returns over months 1–36 after the listing calendar month. The connected portfolio contains firms among the issuer's ten closest text matches that satisfy the six-month pre-offer return requirement. Membership and similarity ranks remain fixed, and monthly portfolio returns equally weight available constituent returns. The SIC2 benchmark averages returns on public firms in the issuer's two-digit industry, excluding the issuer, and requires at least five firms. The resulting panel contains 127,963 issuer-months for 3,942 issuers from the first-day pricing sample. Appendix E describes return availability and delisting conventions.[4]

I estimate

$$R_{i,t}^{IPO} = \alpha_i + \tau_t + \gamma\, R_{i,t}^{Conn} + \delta\, R_{i,t}^{SIC2} + \varepsilon_{i,t}, \tag{9}$$

where i indexes issuers and t indexes calendar months. Issuer fixed effects absorb persistent differences in average returns, and calendar-month fixed effects absorb movements common to issuers in the same month. The coefficient γ measures the issuer's contemporaneous loading

[4]Connected firms must have valid returns on at least 60% of expected trading days in the six-month pre-offer window. Unavailable firms are not replaced or reranked. Issuers remain in the monthly panel through their delisting month, including the delisting return, and leave thereafter. The baseline connected return retains the market-reinvestment convention following constituent delisting; Appendix C.1 also reports returns constructed by dropping connected firms after their delisting month.

on connected returns conditional on its industry return and both fixed effects. Errors are clustered by issuer and calendar month (Petersen, 2009). I report null-imposed wild cluster bootstrap inference for the baseline connected-return coefficient; Appendix F provides the implementation and sensitivity checks.

Connected returns contain information about issuer returns beyond the industry benchmark. In column (1) of Table 15, the connected return coefficient is 0.330 (bootstrap $p < 0.0001$; 95% interval [0.284, 0.376]). The connected portfolio's pooled monthly-return standard deviation is 10.84 percentage points. Multiplying this dispersion by the estimated loading gives a 3.58-percentage-point difference in contemporaneous issuer returns. This scaling uses the pooled return SD before residualizing connected returns on industry returns and the fixed effects.

Table 15: Post-IPO Return Comovement with Connected Firms and Named Competitors

	(1) Baseline	(2) Beyond named competitors	(3) Beyond connected firms
Connected return	0.330***		0.270***
	(0.023)		(0.023)
Connected-only return		0.230***	
		(0.021)	
All named-firm return		0.223***	
		(0.022)	
Named-only return			0.187***
			(0.021)
Bootstrap p-value, connected return	< 0.0001		
SIC2 return	Yes	Yes	Yes
Issuer FE	Yes	Yes	Yes
Calendar-month FE	Yes	Yes	Yes
R^2 net of fixed effects	0.033	0.036	0.035
Issuers	3,942	2,358	2,228
Issuer-months	127,963	76,145	71,840

Notes: The dependent variable is the issuer's monthly return over months 1–36 after listing. Connected returns equally weight available firms among the ten closest text matches; named returns average firms disclosed as competitors. Connected-only and named-only portfolios exclude firms common to both sets. Column (1) reports the baseline; columns (2) and (3) test information beyond named and connected firms, respectively, using each column's available sample. All regressions include SIC2 returns and issuer and calendar-month fixed effects. R^2 removes both fixed effects. Standard errors in parentheses are clustered by issuer and calendar month. For the connected return in column (1), the p-value and stars use a null-imposed wild cluster bootstrap with calendar-month weights and both error-clustering margins in studentization. Other stars use two-way clustered panel-model tests. Appendix F reports confidence intervals, multiple-testing adjustments, and sensitivity checks. Appendix E describes portfolio construction and delisting; Appendix C.1 reports the three-group decomposition and portfolio diagnostics. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

The incremental explanatory power is modest. Adding connected returns raises the R^2 net of issuer and calendar-month fixed effects from 0.026 to 0.033, an increase of approximately 0.0075, or 0.75 percentage points (Appendix Table C.1). The full-model R^2 measures the joint fit of industry and connected returns after removing both fixed effects. Most of the remaining issuer-return variation is unexplained by these benchmarks. These fit measures differ from the average monthly cross-sectional R^2 in Table 3, Panel D, which uses different estimation and return-construction conventions.[5]

The association extends beyond the first year of trading. The connected return coefficients are 0.294 in months 1–12 and 0.339 in months 13–36. Using the respective portfolio-return SDs of 11.45 and 10.48 percentage points gives contemporaneous issuer-return differences of 3.37 and 3.56 percentage points. The original matches therefore remain informative in the second and third years without being updated from subsequent disclosures or issuer returns. The positive relation also holds under alternative return constructions: the coefficient is 0.275 when connected firms are dropped after their delisting month and 0.337 when issuer, connected, and industry monthly returns are winsorized at the 1st and 99th percentiles (Appendix C.1). Weighting by the original cosine scores instead of equally yields an almost identical baseline coefficient, 0.3311 rather than 0.3303 (Appendix Table E.3). Because similarity scores vary little within the selected sets, this comparison represents a modest change in weights.

6.2 Business Descriptions and Named Competitors

I next compare connected firms with public firms named as current or potential competitors in the selected pre-IPO registration statement, including its Business and Risk Factors sections. Text matching measures similarity in reported activities, whereas competitor disclosures record relationships management chooses to name. I examine whether either source contributes return information beyond the other.

I distinguish connected-only firms, which are absent from the recovered named set, from named-only firms, which lie outside the connected set. This separation prevents overlapping firms from supplying the incremental return signal in either comparison. Competitor coverage is incomplete, and eligibility criteria differ between the named and text-based sets

[5]Table 3 uses separate regressions on each comparable-set return, without additional controls or fixed effects, and averages their cross-sectional R^2 across calendar months. Its monthly construction retains partial histories and observed constituent returns without market reinvestment, as described in Appendix E.

(Appendix E). These labels describe membership in the recovered portfolios; connected-only firms may include competitors that were not recovered from the disclosures.

Column (2) of Table 15 includes connected-only returns jointly with returns on all named firms, retaining the industry return and both fixed effects. In the sample of 2,358 issuers and 76,145 issuer-months, the coefficients are 0.230 and 0.223, respectively, both significant at 1%. A one-standard-deviation increase in connected-only returns, 11.27 percentage points, is associated with a 2.59-percentage-point higher issuer return conditional on named-firm returns. Matching business descriptions therefore identifies relevant return variation outside the recovered competitor list.

The converse also holds. Column (3) includes named-only returns alongside returns on all connected firms. The coefficients are 0.187 and 0.270, respectively, both significant at 1%, in a sample of 2,228 issuers and 71,840 issuer-months. A one-standard-deviation increase in named-only returns, 12.98 percentage points, corresponds to a 2.43-percentage-point higher issuer return conditional on connected returns. Explicit competitor disclosures capture relationships beyond the ten closest textual matches. Because columns (2) and (3) use different samples and portfolios, their coefficients and fit do not rank the two selection methods.

A common-sample decomposition provides a further comparison (Appendix Tables C.2 and C.3). Connected-only, named-only, and overlapping portfolios enter jointly, with coefficients of 0.195, 0.212, and 0.144, each significant at 1%. Their respective one-standard-deviation associations are 2.30, 2.59, and 2.53 percentage points. These standardized effects account for differences in return dispersion across the portfolios. This exercise requires all three returns and contains 809 issuers and 26,020 issuer-months. It is a selected subset: 60.2% of issuers in the broader disclosure comparison have no overlap at formation. Within the observed samples, business descriptions and named competitors provide complementary information about issuer return variation.

6.3 Similarity Ranks and Conditional Comovement

The preceding tests assess the selected set as a whole. I next examine whether the similarity ranking distinguishes among related public firms. I partition the thirty highest-ranked firms into three nonoverlapping groups: ranks 1–10, 11–20, and 21–30. Their returns enter equation (9) jointly, together with the industry return and both fixed effects. The ranks are

fixed before applying return-availability requirements. Each coefficient measures a group's loading conditional on the returns of the other groups.

In column (1) of Table 16, the equal-weighted coefficients are 0.277, 0.145, and 0.114 for the three successive groups, all significant at 1%. The sample contains 3,939 issuers and 127,864 issuer-months. The corresponding return SDs are 10.84, 10.62, and 10.40 percentage points, giving one-standard-deviation issuer-return differences of 3.00, 1.54, and 1.18 percentage points. The coefficient on the closest ten firms exceeds that on ranks 11–20 ($p < 0.001$). The positive coefficients on both more distant groups show that related firms outside the closest ten also contain return information.[6]

The distinction between the closest two groups also holds when larger firms receive more weight. Columns (2) and (3) use identical constituents and 127,838 issuer-months for the same 3,939 issuers, require positive lagged market capitalization, and omit months after constituent delisting. The matched equal-weighted coefficients are 0.235, 0.131, and 0.099; the value-weighted coefficients are 0.198, 0.088, and 0.076. Their respective one-standard-deviation associations are 3.03, 1.61, and 1.18 percentage points under equal weighting and 2.40, 1.02, and 0.86 under value weighting. Under both schemes, the coefficient on the closest ten firms exceeds that on the next ten with $p < 0.001$.[7]

Evidence of declines farther down the ranking is weaker. The difference between ranks 11–20 and 21–30 is insignificant in all three columns, with two-sided p-values of 0.190, 0.151, and 0.562. The one-sided tests requiring both adjacent declines yield p-values of 0.095, 0.076, and 0.281. A finer partition into ranks 1–5, 6–10, 11–20, and 21–30 produces coefficients of 0.174, 0.099, 0.145, and 0.114, respectively, all significant at 1%; the second and third groups reverse order (Appendix C.1). Unequal group sizes in this partition also affect portfolio diversification. These supplementary comparisons are unadjusted for multiple testing. They support the clearest distinction at the top of the similarity ranking.

[6]Each rank-band regression requires all three portfolio returns. Column (1) additionally uses the complete-case sample of the finer four-band partition in Appendix C.1. Missing constituent returns do not change ranks or trigger replacement.

[7]The matched equal-weighted return SDs are 12.88, 12.33, and 11.93 percentage points; the value-weighted SDs are 12.11, 11.61, and 11.42. Value weights use the preceding calendar month's market capitalization, normalized within each rank band and month. All individual coefficients in columns (2) and (3) are significant at 1% under the panel-model inference.

Table 16: Business Similarity Rank and Post-IPO Return Comovement

	(1) Baseline equal weights	(2) Matched equal weights	(3) Value weights
Ranks 1–10 return	0.277***	0.235***	0.198***
	(0.019)	(0.015)	(0.016)
Ranks 11–20 return	0.145***	0.131***	0.088***
	(0.021)	(0.017)	(0.014)
Ranks 21–30 return	0.114***	0.099***	0.076***
	(0.019)	(0.014)	(0.016)
SIC2 return	Yes	Yes	Yes
Issuer FE	Yes	Yes	Yes
Calendar-month FE	Yes	Yes	Yes
R^2 net of fixed effects	0.036	0.037	0.033
Issuers	3,939	3,939	3,939
Issuer-months	127,864	127,838	127,838
Coefficient difference tests: two-sided p-values			
Ranks 1–10 versus 11–20	< 0.001	< 0.001	< 0.001
Ranks 11–20 versus 21–30	0.190	0.151	0.562
Both adjacent declines: one-sided p	0.095	0.076	0.281

Notes: The dependent variable is the issuer's monthly return over months 1–36 after listing. Returns for similarity ranks 1–10, 11–20, and 21–30 enter jointly with SIC2 returns and issuer and calendar-month fixed effects. Column (1) equally weights available returns. Columns (2) and (3) use the same constituents and issuer-months, with equal weights and lagged market-capitalization weights, respectively; these columns require positive lagged capitalization and omit months after constituent delisting. R^2 removes both fixed effects. Standard errors are clustered by issuer and calendar month; stars use the corresponding panel-model tests. Difference tests use the joint covariance; the one-sided test requires both adjacent differences to be positive. Tests are unadjusted for multiple comparisons. Appendix E details return construction, and Appendix C.1 reports a finer partition. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

6.4 Business Similarity and Common Economic Exposures

Return comovement may reflect shared cash-flow exposures, discount-rate variation, or correlated investor sentiment. To assess the economic content of the matches, I examine pairwise return correlations after removing conventional risk factors and the issuer's eligible industry benchmark, together with correlations in operating performance.

I pair each issuer with its thirty closest text matches and other eligible public firms in its two-digit SIC industry at the initial-range information cutoff. Equation (14) in Appendix C.3 relates pairwise correlations to standardized pre-IPO cosine similarity, controlling for issuer and public-firm-by-aligned-calendar-window fixed effects, same-SIC2 and same-SIC4 indicators, a named-competitor indicator, and absolute differences in pre-IPO log assets, profitability, and tangibility. Table C.4 also restricts comparisons to the top thirty matches and to firms in the same four-digit SIC industry. Similarity is standardized separately within each estimation sample; differences in coverage and similarity dispersion limit comparisons of coefficient magnitudes across these pools.

For returns, I use daily observations over months 2–12 and 13–36 under the pairwise analysis's convention, which labels the listing calendar month as month 1 and omits it. These windows require at least 120 and 240 common daily observations, respectively; their month labels therefore differ from those of the monthly panel. I adjust both firms' returns for the Fama–French five factors, momentum, and the return of the issuer's fixed eligible SIC2 pool excluding both firms. A one-standard-deviation increase in similarity is associated with increases of 0.0060 and 0.0078 in adjusted return correlation in the two windows, both with $p < 0.001$. These are modest absolute changes in correlations, estimated precisely. The association remains positive within the top thirty matches and within SIC4, indicating that the included factors and industry controls do not fully account for the relation between business similarity and return comovement.[8]

Operating performance provides a complementary outcome measured from firms' accounts. I measure correlations in year-over-year sales growth and year-over-year changes in the quarterly EBITDA-to-assets ratio over the first twelve full calendar quarters after the listing quarter, requiring at least eight aligned observations. A one-standard-deviation increase in

[8]The fully adjusted expanded-pool specifications contain 339,024 pairs for 3,390 issuers in months 2–12 and 259,515 pairs for 3,014 issuers in months 13–36. Factor regressions include an intercept and use common return dates. Pairwise inference uses dyadic standard errors allowing dependence among pairs sharing either firm, including firms appearing in opposite roles (Aronow et al., 2015). The two-sided normal tests are unadjusted for multiple comparisons and remain outside the principal economic-validation family.

similarity is associated with a 0.0249 increase in industry-adjusted sales-growth correlation in the expanded pool ($p < 0.001$; 164,554 pairs and 2,507 issuers). The corresponding sales-growth estimates remain positive in the restricted pools, although the industry-adjusted top-thirty estimate is significant only at 10%. Profitability provides less consistent support: its adjusted relation is insignificant in the expanded pool, positive within the top thirty, and negative within SIC4. Appendix C.3 details alignment, winsorization, and the industry adjustment excluding both firms. The short operating histories and differing samples make these results supporting evidence on common operating exposures.

The sales-growth evidence is consistent with shared demand conditions or other operating exposures contributing to return comovement. The mixed profitability results offer less support for a broader relation across operating fundamentals. Together, the portfolio and pairwise results support the economic relevance of comparables selected from pre-IPO business descriptions, including information beyond broad industry returns and recovered competitor lists.

7 Connected Stock Volatility and IPO Risk

Return comovement indicates whether issuers and their connected firms respond to similar economic developments, but it does not establish whether they face similar levels of equity risk. Businesses exposed to related demand conditions, technologies, and competitive pressures may also face similar sources of uncertainty, which can be reflected in the volatility of their stock returns. For an issuer without a trading history, the pre-offer volatility of these public firms therefore provides a potential benchmark for the risk of its newly listed equity.

This analysis complements evidence linking valuation uncertainty and prospectus language to IPO pricing and subsequent volatility (Lowry et al., 2010; Loughran and McDonald, 2013). I examine whether connected firms' pre-offer total and factor-adjusted volatility predict issuer total volatility across four nonoverlapping windows in the first trading year, beyond industry volatility and observable issuer characteristics.

7.1 Measuring Pre-Offer Connected Risk

I construct connected risk using the public firms identified in Section 3. The measurement window $\mathcal{W}_i$ comprises the six calendar months preceding the offer date and ends on the last trading day strictly before that date. For each connected firm f, total volatility is the

annualized sample standard deviation of raw daily stock returns. Connected total volatility averages these individual firm measures:

$$\mathrm{VOL}_{f,\mathrm{pre}} = \sqrt{252}\,\mathrm{sd}(r_{f,d} : d \in \mathcal{W}_i)\,, \qquad \mathrm{ConnVOL}_{i,\mathrm{pre}} = \frac{1}{K_i}\sum_{f\in\mathcal{C}_i} \mathrm{VOL}_{f,\mathrm{pre}}. \tag{10}$$

Here, $\mathcal{C}_i$ contains eligible firms among the ten closest connected firms, and K_i is their number. The measure captures average firm-level volatility rather than the volatility of a portfolio that diversifies across comparables.

I also measure idiosyncratic volatility (IVOL) using the daily factor model

$$r_{f,d} - r_d^{\mathrm{rf}} = \alpha_f + \beta_f' F_d + \varepsilon_{f,d}, \tag{11}$$

estimated over the same pre-offer window. The factor vector F_d contains the Fama–French five factors and momentum (Fama and French, 2015; Carhart, 1997). Firm-level IVOL is the annualized residual standard error,

$$\widehat{\sigma}^{\mathrm{idio}}_{f,\mathcal{W}_i} = \sqrt{252}\left(\frac{\sum_{d\in\mathcal{W}_i}\widehat{\varepsilon}^2_{f,d}}{n_{f,i}-7}\right)^{1/2},$$

where $n_{f,i}$ is the number of daily observations and seven parameters comprise the intercept and six factor loadings. Connected IVOL is the equal-weighted average of these firm-level estimates.

A firm must have at least 30 valid daily observations and returns on at least 60% of expected trading days. Total volatility retains the same eligible constituents and dates as IVOL. I construct SIC2 industry benchmarks analogously, averaging individual firm measures within the issuer's two-digit industry and excluding the thirty closest connected firms when eligible alternatives remain. The total-volatility and IVOL specifications therefore compare different risk measures using the same constituents and horizon-specific IPO samples.

Idiosyncratic volatility is defined relative to the included return factors; it need not arise from uncertainty unique to a single firm. Herskovic et al. (2016) document substantial common variation in firms' idiosyncratic volatilities even after accounting for common factors in returns. I examine whether business descriptions identify public firms whose residual volatility is informative about a particular issuer.

7.2 Pre-Offer Total Volatility and Issuer Risk

I measure issuer volatility over trading days 2–20, 21–63, 64–126, and 127–252. These nonoverlapping windows distinguish the first weeks of trading from later periods in the first

post-listing year, and all exclude the first-day return. The first window requires at least 15 valid daily returns; the remaining windows require at least 30. For each window h, I estimate

$$\begin{aligned}\log(\text{IPOVOL}_{i,h}) = \alpha_h &+ \beta_h\, z_h(\log \text{ConnVOL}_{i,\text{pre}}) \\ &+ \theta_h\, z_h(\log \text{SIC2VOL}_{i,\text{pre}}) + \Gamma_h' \mathbf{X}_i \\ &+ \mu_{j(i),h} + \tau_{t(i),h} + \xi_{e(i),h} + \varepsilon_{i,h}.\end{aligned} \tag{12}$$

The dependent variable is the log of the sample standard deviation of raw daily issuer returns, annualized by $\sqrt{252}$. Trading days are indexed using valid issuer returns on or after the reconciled first-trading date. The control vector $\mathbf{X}_i$ includes log gross proceeds, $\log(1 + \text{age})$, venture or growth-capital backing, an Internet indicator, and the offer price. All specifications include SIC2 industry, offer-year, and exchange fixed effects. Within each horizon sample, I winsorize log volatility variables at the 1st and 99th percentiles and standardize the signal logs, denoted by $z_h(\cdot)$. Errors are clustered by SIC2 industry and offer month. I report null-imposed wild cluster bootstrap inference for days 21–63 in Panel A; Appendix F provides the implementation and sensitivity checks.

Table 17 reports the SDs used to standardize log connected risk. For total volatility, these SDs range from 0.3914 to 0.3920 across the four windows. A one-SD log increase thus corresponds to approximately 47.9% to 48.0% higher pre-offer connected total volatility.

Issuers whose connected firms are more volatile before the offer exhibit higher total volatility after listing. Panel A of Table 17 reports percentage effects, $100[\exp(\beta_h) - 1]$, per one-standard-deviation increase in log connected volatility. The estimated difference is 12.16% over days 2–20, 9.18% over days 21–63, 7.84% over days 64–126, and 6.79% over days 127–252. For days 21–63, the log coefficient is 0.088 (bootstrap $p = 0.0002$; 95% interval [0.065, 0.112]). The other horizons are also positive and significant at 1%. The corresponding samples contain 3,954, 3,951, 3,938, and 3,903 IPOs.

Median annualized issuer total volatility is 59.44%, 64.98%, 67.86%, and 68.76% in these four estimation samples. Evaluated at the respective medians, Panel A's proportional differences correspond to increases of 7.23, 5.97, 5.32, and 4.67 percentage points in annualized issuer total volatility.

The pre-offer benchmark remains informative from the initial weeks of trading through the second half of the first trading year. Related public firms' trading histories thus reveal differences in issuer equity risk before the issuer begins trading.

Table 17: Pre-Offer Connected Risk and Issuer Total Volatility

Post-listing trading days	2–20 (1)	21–63 (2)	64–126 (3)	127–252 (4)
Panel A: Pre-offer total volatility				
Connected volatility	12.16***	9.18***	7.84***	6.79***
	(1.88)	(1.15)	(1.21)	(1.33)
Bootstrap p-value, connected volatility		0.0002		
SIC2 industry volatility	-2.59	-6.85***	-4.44*	-1.00
	(3.00)	(2.20)	(2.13)	(1.46)
SD of log connected signal	0.3914	0.3915	0.3916	0.3920
N	3,954	3,951	3,938	3,903
Adjusted R^2	0.488	0.550	0.568	0.582
Panel B: Pre-offer idiosyncratic volatility				
Connected IVOL	11.51***	8.86***	7.28***	6.15***
	(1.89)	(1.20)	(1.32)	(1.28)
SIC2 industry IVOL	-0.25	-5.43**	-3.97	-0.57
	(3.31)	(2.31)	(2.39)	(1.63)
SD of log connected signal	0.4066	0.4067	0.4068	0.4068
N	3,954	3,951	3,938	3,903
Adjusted R^2	0.487	0.550	0.567	0.581
Median issuer total volatility (%)	59.44	64.98	67.86	68.76
IPO controls	Yes	Yes	Yes	Yes
SIC2, offer year, exchange FE	Yes	Yes	Yes	Yes

Notes: Each panel reports separate IPO-level OLS regressions of log issuer total volatility over the indicated trading days. Total volatility is the sample standard deviation of raw daily stock returns, annualized by $\sqrt{252}$. Day 1 is excluded; days 2–20 require at least 15 valid returns and later windows require at least 30. Panel A uses pre-offer total volatility; Panel B uses pre-offer FF5 plus momentum residual volatility (IVOL). Connected and SIC2 signals average individual firm volatilities over the six calendar months strictly before the offer date. Connected averages use eligible firms among the ten closest; industry averages exclude the thirty closest firms when eligible alternatives remain. Both panels retain identical constituents and IPO samples within each horizon. Log volatility variables are winsorized at the 1st and 99th percentiles within each horizon sample; signal logs are then standardized. Entries are percentage effects, $100[\exp(\beta) - 1]$, per one-standard-deviation increase in the relevant log signal; parentheses contain delta-method standard errors in percentage units. Controls are log gross proceeds, $\log(1 + \text{age})$, VC or growth-capital backing, an Internet indicator, and offer price. Standard errors cluster by SIC2 and offer month. For connected total volatility over days 21–63 in Panel A, the p-value and stars use a null-imposed wild cluster bootstrap with SIC2 weights and both error-clustering margins in studentization. Other stars use Satterthwaite tests of the log coefficients. Appendix F reports confidence intervals, multiple-testing adjustments, and sensitivity checks. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$. Adjusted R^2 refers to the log-volatility regression. Panels A and B use separate models with different industry risk controls.

The log-signal SD is computed after winsorization within each panel and horizon ($n - 1$ denominator). A one-SD log increase of s corresponds to $100[\exp(s) - 1]\%$ higher underlying connected volatility. Median issuer total volatility is annualized and measured on the corresponding estimation sample, common to both panels. Evaluating a reported percentage effect g at median volatility m in decimal units gives mg percentage points of annualized volatility.

7.3 Pre-Offer Idiosyncratic Volatility and Issuer Total Risk

I next replace connected and SIC2 total volatility in Equation (12) with their IVOL counterparts, retaining issuer total volatility as the outcome and the samples and specification of Panel A. Appendix D reports complementary tests with issuer IVOL as the outcome.

Panel B of Table 17 shows that connected firms' pre-offer idiosyncratic volatility predicts issuer total volatility throughout the first trading year. The SDs of log connected IVOL range from 0.4066 to 0.4068, so a one-SD log increase represents approximately 50.2% higher pre-offer connected IVOL. This increase is associated with 11.51%, 8.86%, 7.28%, and 6.15% higher issuer total volatility across the four successive windows. At the same issuer-volatility medians used for Panel A, these proportional differences correspond to 6.84, 5.76, 4.94, and 4.23 percentage points of annualized issuer total volatility. All four associations are positive and significant.

8 Conclusion

This paper examines what economically similar public firms reveal about firms going public. I identify connected firms by matching business descriptions in pre-IPO registration statements to those in seasoned firms' annual reports using text embeddings. Their leverage, stock returns, and volatility provide information about issuer financing, IPO price formation, and equity risk before the issuer has a public trading history of its own.

Connected leverage provides financing information beyond the issuer's pre-IPO balance sheet. Among issuers initially above connected leverage, those farther above the benchmark reduce more debt relative to pre-IPO assets around the offering, holding starting leverage and conventional financing benchmarks fixed. These findings are consistent with the IPO providing an opportunity to revise relatively high debt positions while persistent financing differences across businesses remain.

Connected returns during bookbuilding predict offer price revisions and remain informative about first-day returns conditional on those revisions and contemporaneous market and industry movements. The first-day relation persists alongside returns on conventional comparables. This pattern is consistent with incomplete offer price adjustment to public information conveyed by connected firms' returns. Selecting comparable firms helps identify information relevant to the difference between the offer price and the first-day closing price,

extending their role beyond assessing valuation levels.

The relevance of these connections extends beyond the offering. Issuer returns comove with connected returns beyond industry movements in both the first year and the following two years after listing. Firms identified from business descriptions and explicitly named competitors provide complementary information about these movements. Because the connections are selected without using the issuer's returns, the comovement evidence gives an economic basis for identifying public comparables from disclosure before a trading history is available.

Connected firms also provide a benchmark for issuer equity risk. Issuers whose connected firms have higher pre-offer total volatility exhibit higher total volatility throughout their first trading year, after accounting for industry volatility and issuer characteristics. The relation also holds when connected and industry risk are measured using factor-adjusted volatility. Related public firms thus provide information about the overall risk of newly listed equity before it can be measured from the issuer's own returns.

Business descriptions do more than summarize an issuer's activities: they identify public firms whose financing choices and market outcomes help characterize the issuer. By linking firms entering public markets to information outside their own records, this approach extends the role of public comparables beyond relative valuation and may also inform the assessment of private firms more generally.

References

Andrew W. Alford. The effect of the set of comparable firms on the accuracy of the price-earnings valuation method. *Journal of Accounting Research*, 30(1):94–108, 1992. doi: 10.2307/2491093.

Aydoğan Alti. How persistent is the impact of market timing on capital structure? *The Journal of Finance*, 61(4):1681–1710, 2006. doi: 10.1111/j.1540-6261.2006.00886.x.

Peter M. Aronow, Cyrus Samii, and Valentina A. Assenova. Cluster-robust variance estimation for dyadic data. *Political Analysis*, 23(4):564–577, 2015. doi: 10.1093/pan/mpv018.

Sanjeev Bhojraj and Charles M. C. Lee. Who is my peer? A valuation-based approach to the selection of comparable firms. *Journal of Accounting Research*, 40(2):407–439, 2002. doi: 10.1111/1475-679X.00054.

Daniel J. Bradley and Bradford D. Jordan. Partial adjustment to public information and IPO underpricing. *Journal of Financial and Quantitative Analysis*, 37(4):595–616, 2002. doi: 10.2307/3595013.

Christian Breitung and Sebastian Müller. Global business networks. *Journal of Financial Economics*, 166:104007, 2025. doi: 10.1016/j.jfineco.2025.104007.

Walid Y. Busaba and Felipe Restrepo. The "7% solution" and IPO (under)pricing. *Journal of Financial Economics*, 144(3):953–971, 2022. doi: 10.1016/j.jfineco.2021.06.041.

A. Colin Cameron, Jonah B. Gelbach, and Douglas L. Miller. Robust inference with multiway clustering. *Journal of Business & Economic Statistics*, 29(2):238–249, 2011. doi: 10.1198/jbes.2010.07136.

Mark M. Carhart. On persistence in mutual fund performance. *The Journal of Finance*, 52 (1):57–82, 1997. doi: 10.1111/j.1540-6261.1997.tb03808.x.

Lauren Cohen and Andrea Frazzini. Economic links and predictable returns. *The Journal of Finance*, 63(4):1977–2011, 2008. doi: 10.1111/j.1540-6261.2008.01379.x.

Francesca Cornelli and David Goldreich. Bookbuilding and strategic allocation. *The Journal of Finance*, 56(6):2337–2369, 2001. doi: 10.1111/0022-1082.00407.

Francesca Cornelli and David Goldreich. Bookbuilding: How informative is the order book? *The Journal of Finance*, 58(4):1415–1443, 2003. doi: 10.1111/1540-6261.00572.

Harry DeAngelo and Richard Roll. How stable are corporate capital structures? *The Journal of Finance*, 70(1):373–418, 2015. doi: 10.1111/jofi.12163.

Roger M. Edelen and Gregory B. Kadlec. Issuer surplus and the partial adjustment of IPO prices to public information. *Journal of Financial Economics*, 77(2):347–373, 2005. doi: 10.1016/j.jfineco.2004.05.009.

Assaf Eisdorfer, Kenneth Froot, Gideon Ozik, and Ronnie Sadka. Competition links and stock returns. *The Review of Financial Studies*, 35(9):4300–4340, 2022. doi: 10.1093/rfs/hhab133.

Eugene F. Fama and Kenneth R. French. A five-factor asset pricing model. *Journal of Financial Economics*, 116(1):1–22, 2015. doi: 10.1016/j.jfineco.2014.10.010.

Michael Faulkender, Mark J. Flannery, Kristine Watson Hankins, and Jason M. Smith. Cash flows and leverage adjustments. *Journal of Financial Economics*, 103(3):632–646, 2012. doi: 10.1016/j.jfineco.2011.10.013.

Thierry Foucault and Laurent Frésard. Learning from peers' stock prices and corporate investment. *Journal of Financial Economics*, 111(3):554–577, 2014. doi: 10.1016/j.jfineco.2013.11.006.

William Grieser, Charles Hadlock, James LeSage, and Morad Zekhnini. Network effects in corporate financial policies. *Journal of Financial Economics*, 144(1):247–272, 2022. doi: 10.1016/j.jfineco.2021.05.060.

Shihao Gu, Bryan Kelly, and Dacheng Xiu. Empirical asset pricing via machine learning. *The Review of Financial Studies*, 33(5):2223–2273, 2020. doi: 10.1093/rfs/hhaa009.

Kathleen Weiss Hanley. The underpricing of initial public offerings and the partial adjustment phenomenon. *Journal of Financial Economics*, 34(2):231–250, 1993. doi: 10.1016/0304-405X(93)90019-8.

Kathleen Weiss Hanley and Gerard Hoberg. The information content of IPO prospectuses. *The Review of Financial Studies*, 23(7):2821–2864, 2010. doi: 10.1093/rfs/hhq024.

Kathleen Weiss Hanley and Gerard Hoberg. Litigation risk, strategic disclosure and the underpricing of initial public offerings. *Journal of Financial Economics*, 103(2):235–254, 2012. doi: 10.1016/j.jfineco.2011.09.006.

Bernard Herskovic, Bryan Kelly, Hanno Lustig, and Stijn Van Nieuwerburgh. The common factor in idiosyncratic volatility: Quantitative asset pricing implications. *Journal of Financial Economics*, 119(2):249–283, 2016. doi: 10.1016/j.jfineco.2015.09.010.

Gerard Hoberg and Gordon Phillips. Product market synergies and competition in mergers and acquisitions: A text-based analysis. *The Review of Financial Studies*, 23(10):3773–3811, 2010. doi: 10.1093/rfs/hhq053.

Gerard Hoberg and Gordon Phillips. Text-based network industries and endogenous product differentiation. *Journal of Political Economy*, 124(5):1423–1465, 2016. doi: 10.1086/688176.

Gerard Hoberg and Gordon M. Phillips. Text-based industry momentum. *Journal of Financial and Quantitative Analysis*, 53(6):2355–2388, 2018. doi: 10.1017/S0022109018000479.

Gerard Hoberg and Gordon M. Phillips. Scope, scale, and concentration: The 21st-century firm. *The Journal of Finance*, 80(1):415–466, 2025. doi: 10.1111/jofi.13400.

Manish Jha, Hongyi Liu, and Asaf Manela. Does finance benefit society? A language embedding approach. *The Review of Financial Studies*, 39(5):1227–1266, 2026. doi: 10.1093/rfs/hhaf012.

Moonchul Kim and Jay R. Ritter. Valuing IPOs. *Journal of Financial Economics*, 53(3): 409–437, 1999. doi: 10.1016/S0304-405X(99)00027-6.

Yongjin Kim, Lars-Alexander Kuehn, and Kai Li. Learning about the consumption risk exposure of firms. *Journal of Financial Economics*, 152:103759, 2024. doi: 10.1016/j.jfineco.2023.103759.

Mark T. Leary and Michael R. Roberts. Do firms rebalance their capital structures? *The Journal of Finance*, 60(6):2575–2619, 2005. doi: 10.1111/j.1540-6261.2005.00811.x.

Mark T. Leary and Michael R. Roberts. Do peer firms affect corporate financial policy? *The Journal of Finance*, 69(1):139–178, 2014. doi: 10.1111/jofi.12094.

Charles M. C. Lee, Paul Ma, and Charles C. Y. Wang. Search-based peer firms: Aggregating investor perceptions through internet co-searches. *Journal of Financial Economics*, 116(2): 410–431, 2015. doi: 10.1016/j.jfineco.2015.02.003.

Charles M. C. Lee, Terrence Tianshuo Shi, Stephen Teng Sun, and Ran Zhang. Production complementarity and information transmission across industries. *Journal of Financial Economics*, 155:103812, 2024. doi: 10.1016/j.jfineco.2024.103812.

Michael L. Lemmon, Michael R. Roberts, and Jaime F. Zender. Back to the beginning: Persistence and the cross-section of corporate capital structure. *The Journal of Finance*, 63(4):1575–1608, 2008. doi: 10.1111/j.1540-6261.2008.01369.x.

Kai Li, Feng Mai, Rui Shen, and Xinyan Yan. Measuring corporate culture using machine learning. *The Review of Financial Studies*, 34(7):3265–3315, 2021. doi: 10.1093/rfs/hhaa079.

Kai Li, Feng Mai, Rui Shen, Chelsea Yang, and Tengfei Zhang. Dissecting corporate culture using generative AI. *The Review of Financial Studies*, 39(1):253–296, 2026. doi: 10.1093/rfs/hhaf081.

Jing Liu, Doron Nissim, and Jacob Thomas. Equity valuation using multiples. *Journal of Accounting Research*, 40(1):135–172, 2002. doi: 10.1111/1475-679X.00042.

Tim Loughran and Bill McDonald. IPO first-day returns, offer price revisions, volatility, and form S-1 language. *Journal of Financial Economics*, 109(2):307–326, 2013. doi: 10.1016/j.jfineco.2013.02.017.

Michelle Lowry and G. William Schwert. Is the IPO pricing process efficient? *Journal of Financial Economics*, 71(1):3–26, 2004. doi: 10.1016/S0304-405X(03)00205-8.

Michelle Lowry, Micah S. Officer, and G. William Schwert. The variability of IPO initial returns. *The Journal of Finance*, 65(2):425–465, 2010. doi: 10.1111/j.1540-6261.2009.01540.x.

Michelle Lowry, Roni Michaely, and Ekaterina Volkova. Information revealed through the regulatory process: Interactions between the SEC and companies ahead of their IPO. *The Review of Financial Studies*, 33(12):5510–5554, 2020. doi: 10.1093/rfs/hhaa007.

Peter MacKay and Gordon M. Phillips. How does industry affect firm financial structure? *The Review of Financial Studies*, 18(4):1433–1466, 2005. doi: 10.1093/rfs/hhi032.

Lior Menzly and Oguzhan Ozbas. Market segmentation and cross-predictability of returns. *The Journal of Finance*, 65(4):1555–1580, 2010. doi: 10.1111/j.1540-6261.2010.01578.x.

Marco Pagano, Fabio Panetta, and Luigi Zingales. Why do companies go public? An empirical analysis. *The Journal of Finance*, 53(1):27–64, 1998. doi: 10.1111/0022-1082.25448.

Mitchell A. Petersen. Estimating standard errors in finance panel data sets: Comparing approaches. *The Review of Financial Studies*, 22(1):435–480, 2009. doi: 10.1093/rfs/hhn053.

Amiyatosh K. Purnanandam and Bhaskaran Swaminathan. Are IPOs really underpriced? *The Review of Financial Studies*, 17(3):811–848, 2004. doi: 10.1093/rfs/hhg055.

David Roodman, Morten Ørregaard Nielsen, James G. MacKinnon, and Matthew D. Webb. Fast and wild: Bootstrap inference in stata using boottest. *The Stata Journal*, 19(1):4–60, 2019. doi: 10.1177/1536867X19830877.

Suproteem K. Sarkar and Keyon Vafa. Lookahead bias in pretrained language models. Working paper 4754678, SSRN, 2024. doi: 10.2139/ssrn.4754678.

Tyler Shumway. The delisting bias in CRSP data. *Journal of Finance*, 52(1):327–340, 1997.

Tyler Shumway and Vincent A. Warther. The delisting bias in CRSP's Nasdaq data and its implications for the size effect. *Journal of Finance*, 54(6):2361–2379, 1999.

Matthew Spiegel and Heather Tookes. Why does an IPO affect rival firms? *The Review of Financial Studies*, 33(7):3205–3249, 2020. doi: 10.1093/rfs/hhz081.

A Connected Firm Leverage

A.1 Issuer Characteristics and Aggregation

Adding issuer pre-IPO profitability, size, and tangibility on identical samples reduces the connected leverage coefficient from 0.158 to 0.129 for book leverage and from 0.225 to 0.174 for market leverage. After adding these characteristics, book leverage is significant at 10% and market leverage at 5% under the preferred fixed effects. With SIC2-by-three-year-period fixed effects, both estimates are significant at 10% under Satterthwaite inference.

Table A.1 varies the number of connected firms and their portfolio weights. Market capitalization weights apply to both leverage and connected firm controls. Equal-weighted book leverage coefficients are positive and significant at 5% across connected set sizes except $K = 10$, which reaches 10%. No value-weighted book leverage coefficient is significant at 10%. Equal-weighted market leverage reaches 10% at $K = 5$ and 5% at larger set sizes. Value-weighted market leverage reaches 10% at $K = 10$ and 5% for $K \geq 15$; the $K = 5$ estimate is not significant at 10%.

Table A.1: Connected Firm Leverage: Number of Connected Firms and Weighting

	$K = 5$	$K = 10$	$K = 15$	$K = 20$	$K = 25$	$K = 30$
Panel A: IPO book leverage						
Equal-weighted	0.116**	0.129*	0.169**	0.198**	0.231**	0.259**
	(0.044)	(0.055)	(0.063)	(0.078)	(0.077)	(0.089)
Value-weighted	0.072	0.044	0.011	0.017	0.053	0.048
	(0.046)	(0.033)	(0.028)	(0.033)	(0.027)	(0.033)
N (equal-weighted)	3,727	3,728	3,728	3,728	3,728	3,728
Panel B: IPO market leverage						
Equal-weighted	0.109*	0.174**	0.252**	0.279**	0.287**	0.317**
	(0.055)	(0.067)	(0.081)	(0.095)	(0.100)	(0.106)
Value-weighted	0.101	0.117*	0.153**	0.169**	0.201**	0.192**
	(0.054)	(0.057)	(0.054)	(0.066)	(0.063)	(0.070)
N (equal-weighted)	3,725	3,726	3,726	3,726	3,726	3,726

Notes: Each cell reports the connected firm leverage coefficient from a separate regression of issuer book leverage (Panel A) or market leverage (Panel B) at listing. Columns vary the number of connected firms, K; rows use equal or CRSP market capitalization weights, with capitalization measured near the accounting date. The same weights apply to connected firm leverage and controls. All regressions include the SIC2 industry benchmark, IPO controls, issuer pre-IPO characteristics, connected firm controls, and offer year, SIC2, and exchange fixed effects, as in column (4) of Table 10. Winsorization follows that table. Reported N refers to equal weighting; value-weighted samples differ slightly. Appendix A discusses the weighting comparison. CR2 standard errors in parentheses are two-way clustered by issuer SIC2 and offer year. Stars use Satterthwaite p-values: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

A.2 Industry and Alternative Comparisons

Column (5) of Table 10 drops the SIC2 fixed effects while retaining all controls in column (4). Without industry fixed effects, the connected and industry leverage coefficients are significant at 5% for both outcomes. With SIC2 fixed effects, the industry coefficients are not significant at 10%.

Table 11 compares connected leverage with SIC2 industry leverage and with the leverage of five firms matched across industries on pre-IPO log assets, profitability, and tangibility. The comparisons standardize each leverage measure within the full book- and market-leverage samples of 3,728 and 3,726 issuers. Connected market leverage is significant at 5% under both inference procedures in both comparisons, while connected book leverage is significant at 10%. Characteristic-matched leverage remains significant for both outcomes; the SIC2 coefficients are not significant at 10% conditional on the controls and fixed effects. None of the connected-minus-alternative coefficient differences is significant at 5%.

A.3 Inherited Versus Post-IPO Leverage

Table A.2 repeats Table 12 under alternative samples and specifications. Panel A restricts every column to the 3,148 issuers observed at all four dates. The one- and two-year conditional coefficients are 0.169 and 0.219. The former is significant at 5% under both procedures and the latter at 10%; the listing-quarter coefficient is not significant at 5%. Panel B uses fixed effects for SIC2 interacted with three-year periods. Its two-year coefficient is 0.268, with Satterthwaite and wild bootstrap p-values of 0.052 and 0.074. Panel C caps the issuer's pre-IPO leverage ratio at one to limit the influence of issuers reporting debt above book assets. The one- and two-year coefficients are 0.145 and 0.188, with wild cluster bootstrap p-values of 0.074 and 0.109.

Table A.2: Issuer Leverage Before and After the IPO: Alternative Samples and Specifications

Issuer book leverage measured at:	Pre-IPO (1)	Listing (2)	+1 year (3)	+2 years (4)
Panel A: Balanced sample, all four outcomes observed				
Without issuer pre-IPO leverage				
Connected firm book leverage	0.309**	0.133*	0.215**	0.267**
	(0.113)	(0.064)	(0.069)	(0.101)
Wild-cluster p-value	[0.040]	[0.079]	[0.027]	[0.050]
Observations	3,148	3,148	3,148	3,148
Conditional on issuer pre-IPO leverage				
Connected firm book leverage		0.080	0.169**	0.219*
		(0.055)	(0.068)	(0.102)
Wild-cluster p-value		[0.172]	[0.049]	[0.088]
Issuer pre-IPO book leverage		0.171***	0.149***	0.154***
		(0.024)	(0.014)	(0.014)
Observations		3,148	3,148	3,148
Panel B: SIC2-by-three-year-period fixed effects				
Without issuer pre-IPO leverage				
Connected firm book leverage	0.353***	0.143*	0.232**	0.313**
	(0.070)	(0.057)	(0.084)	(0.104)
Wild-cluster p-value	[0.012]	[0.061]	[0.058]	[0.054]
Observations	3,699	3,699	3,514	3,160
Conditional on issuer pre-IPO leverage				
Connected firm book leverage		0.085	0.179*	0.268*
		(0.046)	(0.085)	(0.104)
Wild-cluster p-value		[0.131]	[0.114]	[0.074]
Issuer pre-IPO book leverage		0.163***	0.150***	0.150***
		(0.021)	(0.013)	(0.014)
Observations		3,699	3,514	3,160
Panel C: Issuer pre-IPO leverage capped at one				
Without issuer pre-IPO leverage				
Connected firm book leverage	0.289***	0.125*	0.229**	0.263**
	(0.071)	(0.056)	(0.071)	(0.100)
Wild-cluster p-value	[0.017]	[0.066]	[0.031]	[0.048]
Observations	3,699	3,699	3,514	3,160
Conditional on issuer pre-IPO leverage				
Connected firm book leverage		0.028	0.145*	0.188*
		(0.040)	(0.064)	(0.094)
Wild-cluster p-value		[0.488]	[0.074]	[0.109]
Issuer pre-IPO book leverage		0.335***	0.295***	0.296***
		(0.023)	(0.024)	(0.018)
Observations		3,699	3,514	3,160

Notes: Each panel repeats Table 12 under an alternative sample or specification. Panel A restricts every column to issuers for which all four leverage measures and all regressors are observed. Panel B replaces the SIC2 fixed effects with SIC2-by-three-year-period fixed effects, retaining offer year and exchange fixed effects. Panel C caps the issuer's winsorized pre-IPO book leverage at one, both as the outcome in column (1) and as the control in the conditional regressions; 7% of issuers report pre-IPO debt above book assets. Outcomes, controls, and inference follow Table 12. CR2 standard errors in parentheses are two-way clustered by issuer SIC2 and offer year. Stars use Satterthwaite p-values: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$. Brackets report wild-cluster bootstrap p-values for connected firm leverage (4,999 draws; the same cluster dimensions).

A.4 Debt Changes Around the IPO

Because pre-IPO debt enters the outcome with a negative sign, the coefficient on initial leverage is not a speed of adjustment or a fraction of a leverage gap closed. I estimate

$$\begin{aligned}\frac{D_{i,\text{after}} - D_{i,\text{before}}}{A_{i,\text{before}}} = \alpha + \beta\, \text{ConnLev}_{i,t-1}^{\text{book},K=10} + \phi\, \frac{D_{i,\text{before}}}{A_{i,\text{before}}} \\ + \theta\, \text{IndLev}_{j,t-1}^{\text{book}} + \Gamma' \overline{\mathbf{Z}}_{i,t-1}^{K=10} + \Lambda' \mathbf{X}_{it} \\ + \Psi' \mathbf{W}_{i,t-1} + \mu_j + \tau_t + \varepsilon_{ijt}.\end{aligned} \tag{13}$$

Here $D_{i,\text{before}}$ and $D_{i,\text{after}}$ are total debt, measured as current plus long-term debt, at the latest quarter-end strictly before the offer and the first quarter-end on or after it, respectively, with the post-IPO endpoint within 270 days of the offer. $A_{i,\text{before}}$ is book assets at the pre-IPO quarter-end. The ratio $D_{i,\text{before}}/A_{i,\text{before}}$ controls for the issuer's starting leverage. The coefficient β measures the difference in debt changes, scaled by pre-IPO assets, per unit of connected book leverage, holding the other regressors fixed.

Indices follow Equation (7), and $\text{IndLev}^{\text{book}}$ is average book leverage among public firms in the same SIC2 industry. The vectors $\overline{\mathbf{Z}}$, $\mathbf{X}$, and $\mathbf{W}$ contain the connected firm characteristics, IPO controls (including exchange fixed effects), and issuer pre-IPO financial characteristics defined in Equation (7). Issuer profitability, size, and tangibility use the latest annual statement dated before the offer. The terms μ_j and τ_t denote SIC2 and offer year fixed effects. Column (2) of Table A.3 replaces μ_j with fixed effects for SIC2 interacted with three-year periods and retains offer year and exchange fixed effects.

The connected-leverage coefficient in Table A.3 is not significant at 5% conditional on initial leverage, industry leverage, and issuer and connected firm characteristics. The coefficients are 0.173 in column (1) and 0.210 with SIC2-by-three-year-period fixed effects in column (2), with wild cluster bootstrap p-values of 0.099 and 0.081, respectively.

Table A.4 reports tighter accounting windows, complete debt reporting, and uncapped ratio checks, each including issuer pre-IPO profitability, size, and tangibility. Neither the tighter-window nor the complete-reporting coefficient reaches 5% under Satterthwaite inference in either specification.

Without winsorizing the debt-change and initial-leverage ratios, the main coefficient rises to 0.319 (wild cluster bootstrap $p = 0.082$), and the stricter fixed-effect estimate is 0.496 ($p = 0.027$). The stricter specification therefore reaches 5% only when these ratios are left

Table A.3: Connected Firm Leverage and Debt Changes Around the IPO

	(1) Main	(2) Stricter industry/time
Connected firm book leverage	0.173*	0.210*
	(0.082)	(0.087)
Issuer pre-IPO book leverage	-0.646***	-0.646***
	(0.028)	(0.028)
Wild cluster bootstrap p-value	[0.099]	[0.081]
Issuer profitability, size, and tangibility	Yes	Yes
Other controls and SIC2 industry benchmark	Yes	Yes
Offer-year and exchange fixed effects	Yes	Yes
Industry fixed effects	SIC2	SIC2 × three-year period
N	3,699	3,699
Adjusted R^2	0.626	0.624
Mean debt change/pre-IPO assets	-0.152	-0.152
Median debt change/pre-IPO assets	-0.018	-0.018
Issuers with a debt decline (%)	65.8	65.8

Notes: The outcome is the change in current plus long-term debt around the IPO, divided by pre-IPO assets. Connected firm book leverage averages the ten closest connected firms equally. Both regressions control for issuer pre-IPO book leverage; other controls follow column (4) of Table 10. Column (1) includes SIC2, offer year, and exchange fixed effects; column (2) replaces SIC2 fixed effects with SIC2-by-three-year-period fixed effects. Appendix A.4 defines the accounting endpoints; Appendix E describes construction and winsorization. Descriptive statistics use the estimation sample and winsorized outcome. CR2 standard errors in parentheses are two-way clustered by issuer SIC2 and offer year. Stars use Satterthwaite p-values: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$. Brackets report wild cluster bootstrap p-values for connected firm book leverage (4,939 and 4,920 usable draws in columns (1) and (2), respectively; the same cluster dimensions).

unwinsorized. The outcome throughout is a change in balance-sheet debt outstanding, not issuance and repayment cash flows.

A.5 Linear Initial Leverage and Asymmetric Financing Changes

Table A.5 estimates Equation (8) with linear initial leverage. Among issuers starting above the connected benchmark, a one-standard-deviation lower benchmark is associated with debt changes lower by 4.58% of pre-IPO assets, holding initial leverage and the remaining controls fixed. The common-slope null $\beta^A + \beta^B = 0$ is rejected ($p = 0.004$, Satterthwaite); the below-benchmark slope is not significant at 5%. The main-text specification allows initial leverage to enter flexibly and adds competing financing benchmarks.

The leverage change estimates with linear initial leverage show the same asymmetry at

Table A.4: Debt Change Regressions: Issuer Controls and Robustness

	Main		Stricter industry/time	
	Coefficient	N	Coefficient	N
Panel A: Same-sample comparison				
Without the three issuer characteristics	0.190*	3,699	0.225*	3,699
	(0.090)		(0.097)	
With the three issuer characteristics	0.173*	3,699	0.210*	3,699
	(0.082)		(0.087)	
Panel B: Expanded-control sensitivity checks				
Tighter accounting window	0.174*	3,666	0.213*	3,666
	(0.080)		(0.085)	
Complete debt reporting	0.155*	3,588	0.184*	3,588
	(0.079)		(0.084)	
Uncapped debt change and starting-leverage ratios	0.319	3,699	0.496**	3,699
	(0.167)		(0.172)	

Notes: Each cell reports the connected firm book leverage coefficient for debt change/pre-IPO assets, with controls and fixed effects as in Table A.3. Panel A holds the sample fixed when adding issuer pre-IPO profitability, log assets, and tangibility; Panel B includes all three. The tighter window limits the accounting span to 180 days and requires a post-offer quarter-end. Complete reporting requires both debt components at both endpoints. The uncapped check leaves debt change and starting leverage unwinsorized. Other variables retain baseline treatment, and restricted samples retain the original cutoffs (Appendix E). CR2 standard errors in parentheses are two-way clustered by issuer SIC2 and offer year. Stars use Satterthwaite p-values: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

listing and one and two years later, with the common-slope null rejected at every horizon. The above-benchmark debt-change coefficient remains negative at longer horizons, but the common-slope null for debt changes is not rejected at 5% after either one or two years.

A.6 Alternative Financing Benchmarks

Table A.6 repeats Equation (8) with the leverage gap defined using connected firms, the SIC2 industry average, or characteristic-matched firms. Both gap components enter each regression, although Panel A reports the above-benchmark slopes. This comparison retains ten firms selected using the matching characteristics of Table 11. All columns omit the SIC2-level control because, in the SIC2 specification, that variable equals $\text{PreLev} + \text{BelowGap}_{SIC2} - \text{AboveGap}_{SIC2}$; the remaining controls and fixed effects are unchanged. The comparison retains the common sample with coverage for all benchmarks, including the requirement that

Table A.5: Leverage Gaps and Issuer Financing Changes

	Change in book leverage			Debt change/pre-IPO assets		
	Listing (1)	+1 year (2)	+2 years (3)	Listing (4)	+1 year (5)	+2 years (6)
Below-benchmark distance (BelowGap)	-0.170***	-0.041	0.003	-0.127*	0.697*	0.598
	(0.041)	(0.065)	(0.090)	(0.062)	(0.310)	(0.354)
Wild cluster bootstrap p-value	[0.006]	[0.546]	[0.980]	[0.060]	[0.077]	[0.145]
Above-benchmark distance (AboveGap)	-0.231***	-0.329***	-0.368***	-0.388**	-1.480***	-1.420*
	(0.053)	(0.067)	(0.090)	(0.110)	(0.407)	(0.705)
Wild cluster bootstrap p-value	[0.009]	[0.014]	[0.014]	[0.030]	[0.028]	[0.127]
Issuer pre-IPO book leverage	-0.631***	-0.545***	-0.506***	-0.297**	0.821*	0.941
	(0.053)	(0.065)	(0.091)	(0.100)	(0.379)	(0.679)
$H_0: \beta^A + \beta^B = 0$: p-value	0.0005	0.0007	0.0007	0.0044	0.0984	0.2746
N	3,699	3,514	3,160	3,699	3,514	3,160
R^2	0.890	0.880	0.820	0.643	0.191	0.141

Notes: The nonnegative distances are AboveGap = max(PreLev − ConnLev, 0) and BelowGap = max(ConnLev − PreLev, 0). Both enter each regression. The common-slope null is $\beta^A + \beta^B = 0$. Connected leverage averages the ten closest connected firms equally. The outcomes are the change in book leverage and the change in debt divided by pre-IPO assets, from the last quarter-end before the offer to the indicated horizon. Accounting dates follow Table 12. All columns include initial leverage, SIC2 industry leverage, issuer and connected firm characteristics, IPO controls, and SIC2, offer year, and exchange fixed effects. Leverage changes and gaps subtract existing winsorized ratios without further winsorization; debt changes/pre-IPO assets are winsorized at the 1st and 99th percentiles. CR2 standard errors in parentheses are two-way clustered by SIC2 and offer year. Stars and the common-slope test use Satterthwaite p-values: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$. Brackets report wild cluster bootstrap p-values (4,383–4,992 usable draws, depending on the test).

SIC4 leverage be observed, and does not recompute winsorization cutoffs on this smaller sample.

Panel B reports the difference between the connected-firm and alternative above-benchmark slopes. These are comparisons of separate specifications, with every control and fixed-effect coefficient allowed to differ across equations. The paired wild cluster bootstrap imposes equality of the slopes and preserves the dependence between estimates by assigning the same SIC2 and offer-year cluster weights to both observations of each issuer. The reported p-values

are two-sided and are not adjusted for multiple testing.

Table A.6: Asymmetric Financing Changes: Alternative Benchmarks

	Change in book leverage			Debt change/pre-IPO assets		
	Listing	+1 year	+2 years	Listing	+1 year	+2 years
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Above-benchmark slopes						
Connected firms	-0.268***	-0.370***	-0.389***	-0.460***	-1.755***	-1.646**
	(0.053)	(0.060)	(0.084)	(0.106)	(0.398)	(0.669)
SIC2 industry	-0.196	-0.153	-0.161	-0.461	-0.748	-0.725
	(0.172)	(0.187)	(0.214)	(0.268)	(0.831)	(0.922)
Characteristic-matched firms	-0.196***	-0.222***	-0.210**	-0.156	-0.426	-0.072
	(0.035)	(0.036)	(0.055)	(0.095)	(0.443)	(0.669)
N	3,171	3,015	2,699	3,171	3,015	2,699
Panel B: Connected minus alternative slope						
Connected minus SIC2	-0.072	-0.217	-0.228	0.001	-1.007	-0.921
	(0.151)	(0.221)	(0.239)	(0.187)	(0.966)	(1.262)
Paired wild p-value	0.5863	0.3012	0.3338	0.9946	0.3048	0.4630
Connected minus characteristic matched	-0.072	-0.148	-0.179	-0.305**	-1.329*	-1.574
	(0.046)	(0.077)	(0.087)	(0.090)	(0.667)	(0.985)
Paired wild p-value	0.1759	0.1301	0.1091	0.0169	0.0586	0.1498

Notes: Each Panel A coefficient is from a separate regression containing both gap components and initial leverage, with outcomes, remaining controls, and fixed effects as in Table A.5. Nonnegative distances follow Table A.5. The SIC2-level control is omitted throughout because industry leverage equals initial leverage plus the below distance minus the above distance. Characteristic matching selects ten public firms across industries by pre-IPO log assets, profitability, and tangibility, using the same matching characteristics as Table 11. The common sample requires coverage for connected, SIC2, SIC4, and characteristic-matched leverage; the three benchmarks shown use identical IPOs within each horizon. Existing transformations are retained. Panel B tests differences between the separate-equation slopes, retaining dependence across estimates through shared SIC2 and offer-year cluster weights. Standard errors are two-way CR2. Panel A stars use Satterthwaite p-values; Panel B stars and displayed p-values use the paired wild bootstrap (3,856–4,999 usable draws, depending on the test). All tests are two-sided without a multiple-testing adjustment: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

For debt changes at listing, the connected above-benchmark slope is more negative than the characteristic-matched slope (paired wild cluster bootstrap $p = 0.017$). The corresponding leverage-change comparisons are not significant at 5% at any horizon. None of the comparisons with SIC2 leverage rejects equal slopes.

A.7 Predicted Leverage and Additional Financing Controls

The predicted leverage benchmark in Table 13 summarizes the relation between book leverage and conventional firm characteristics among public firms. The candidate firms are those in the preceding fiscal-year cross-section. Each contributes its latest eligible annual record with a verified filing date strictly before the offer, using an older period when necessary and never a period later than the original record. Firms without a verified pre-offer record are omitted from model estimation. Firms must have an active CRSP link at the statement date and observed book leverage, log assets, operating profitability, tangibility, and SIC2 classification. Financial firms, utilities, and the issuer itself are excluded. No missing input or disclosure date is imputed.

Filing dates come from WRDS company filing records and SEC submissions histories matched to the firm and reporting period. The cutoff applies to the observations used for winsorization, industry means, and model estimation. The regressions use current Compustat values. I interpret the estimates as retrospective conditional comparisons.

Book leverage is capped at one. Book leverage and the firm characteristics are winsorized at the 1st and 99th percentiles of the eligible estimation sample; the same characteristic cutoffs are applied to the issuer. The industry regressor is the mean book leverage of the other eligible firms in the same SIC2 industry. Firms in singleton industries cannot supply this regressor and are excluded from the fit. For the IPO prediction, industry leverage is the mean among eligible public firms in its industry, excluding the issuer. The OLS fitted value is retained without clipping. Two IPOs lack an eligible industry input, leaving 3,697 observations at listing, 3,512 after one year, and 3,158 after two years. Panel A uses the same observations as Panel B at each horizon. The inference in both panels conditions on the fitted benchmark and does not incorporate first-stage estimation uncertainty. The fitted value is a conditional leverage benchmark.

Table A.7 adds the change between the issuer's last pre-IPO leverage observation and the quarter-end exactly three months earlier. Both source dates must have unique records, and the prior leverage ratio must be observed. The recent-change control is winsorized at the observed IPO sample's 1st and 99th percentiles. All four columns use the resulting 2,523 issuers; columns (1) and (3) omit this control to hold the sample fixed. Among issuers starting above the connected benchmark, the estimated additional debt reduction for a one-standard-deviation lower benchmark is 2.68% of pre-IPO assets without the recent-change

control and 2.64% with it. The wild cluster bootstrap p-value is 0.070 in both cases. Neither listing leverage coefficient reaches 5%.

Table A.7: Predicted Leverage and Recent Pre-IPO Leverage Changes

	Change in book leverage		Debt change/A_{pre}	
	Baseline	+Recent change	Baseline	+Recent change
	(1)	(2)	(3)	(4)
Connected below-benchmark distance (BelowGap_C)	-0.063	-0.061	0.032	0.032
	(0.034)	(0.033)	(0.053)	(0.052)
Wild p-value (BelowGap_C)	0.103	0.099	0.592	0.575
Connected above-benchmark distance (AboveGap_C)	-0.085	-0.079	-0.227*	-0.224*
	(0.062)	(0.061)	(0.098)	(0.097)
Wild p-value (AboveGap_C)	0.241	0.266	0.070	0.070
Predicted below-benchmark distance (BelowGap_P)	-0.080	-0.084	-0.106	-0.109
	(0.075)	(0.069)	(0.206)	(0.209)
Wild p-value (BelowGap_P)	0.311	0.256	0.634	0.639
Predicted above-benchmark distance (AboveGap_P)	-0.271*	-0.276*	-0.337	-0.339
	(0.136)	(0.140)	(0.229)	(0.231)
Wild p-value (AboveGap_P)	0.063	0.063	0.166	0.169
Recent pre-IPO leverage change		-0.043		-0.022
		(0.022)		(0.028)
Wild p-value (recent change)		0.112		0.625
N	2,523	2,523	2,523	2,523
R^2	0.914	0.914	0.714	0.714

Notes: All columns use the same 2,523 IPOs with unique records in both quarters. The first two columns use the listing book leverage change; the last two use the listing debt change divided by pre-IPO assets. All regressions include connected, industry, and predicted leverage gaps, the initial leverage spline, and the remaining controls and fixed effects of Table 13, Panel B. Columns (2) and (4) add the change from the quarter-end three months earlier to the last pre-IPO quarter-end, winsorized at the observed IPO sample's 1st and 99th percentiles. Nonnegative above- and below-benchmark distances follow Table 13. Inference conditions on the fitted benchmark. Parentheses report CR2 standard errors clustered by SIC2 and offer year. Stars use Satterthwaite p-values: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$. Wild cluster bootstrap p-values use 4,898–4,997 usable draws, depending on the test. Tests are two-sided without a multiple-testing adjustment.

A.8 Observed Pre-IPO Leverage Histories

Table A.8 reports historical coverage and leverage positions for issuers above connected leverage in their last pre-IPO quarter. The connected benchmark is held fixed at its IPO value to assess whether the observed excess predates that quarter. Requiring both debt components, positive assets and matching currencies preserves observed accounting histories without filling missing quarters.

Table A.8: Observed Pre-IPO Histories of Issuers Above Connected Leverage

Earlier quarter	Matched IPOs	Coverage (%)	Already above benchmark (%)	Earlier median excess (pp)	Median excess at quarter −1 (pp)
−12	60	2.8	88.3	27.9	36.8
−11	72	3.4	87.5	31.1	36.8
−10	102	4.8	90.2	28.9	35.6
−9	226	10.6	73.9	23.0	35.1
−8	411	19.3	78.3	20.6	29.3
−7	431	20.2	76.3	20.6	26.0
−6	544	25.5	78.9	21.8	27.7
−5	442	20.7	80.3	23.2	31.1
−4	636	29.8	85.2	25.7	29.8
−3	624	29.2	87.7	25.2	28.0
−2	1,490	69.8	91.6	27.1	29.6

Notes: The parent sample contains 3,697 IPOs with the connected, industry and predicted leverage inputs of Table 13. Issuers are classified using their existing winsorized pre-IPO and connected leverage measures, and the connected benchmark remains fixed at its IPO value. Quarter 0 contains the offering; quarter −1 is the last pre-offer fiscal quarter-end. Earlier dates use exact three-month fiscal offsets, without substituting missing quarters. Each row requires observed leverage at both the earlier date and quarter −1; the two excess medians therefore use identical IPOs within a row. Historical leverage uses both reported debt components, positive assets and a currency matching the pre-IPO anchor. Excess is raw issuer leverage minus the fixed connected benchmark. Coverage is the fraction of the 2,135 initially above-benchmark issuers with observed quarter −1 leverage. Rows use different, selectively observed historical samples.

A.9 Debt Reductions Relative to Initial Balances

A complementary outcome measures debt reductions relative to the issuer's initial debt balance:

$$\text{ReductionFraction}_i = \frac{\max(D_{i,\text{pre}} - D_{i,\text{listing}}, 0)}{D_{i,\text{pre}}}.$$

The outcome is zero for issuers whose debt does not fall, capturing both the occurrence and size of reductions. I require initial debt of at least 5% of assets and retain the financing controls in Table 13, Panel B. Both debt components must be reported at each endpoint, with matching currencies. In the 2,602-IPO sample, a one-standard-deviation lower connected benchmark is associated with a 2.46-percentage-point larger fraction of initial debt reduced

for issuers starting above the benchmark (Satterthwaite $p = 0.073$; wild bootstrap $p = 0.067$; Appendix Table A.9). The estimates remain positive but become less precise when the initial-debt threshold is raised or the sample excludes extreme initial leverage. This measure scales the adjustment by the debt balance the issuer brings to the offering.

Table A.9 reports alternative initial-debt thresholds. The above-benchmark gap coefficient for the fraction of initial debt reduced with zeros retained is significant at 10% with the lower debt threshold, but not at 5%. The probability, signed-fraction, prospectus-intention, and primary-proceeds interaction estimates are also not significant at 5%. Samples and inference differ across these outcomes as reported in the table. These outcomes are distinct from the asset-scaled debt changes in Table 13.

The repayment-plan indicator records an affirmative allocation of IPO proceeds to existing debt principal in a readable pre-offer Use of Proceeds section. No explicit allocation is coded as zero only when the relevant text is available. These regressions require observed initial debt and USD-denominated initial assets and add gross proceeds divided by initial assets; complete listing debt is required only for realized balance outcomes. Funding regressions use USD-denominated base-offering data, with primary and secondary gross proceeds constructed from share counts and the offer price, before fees and excluding over-allotments. Both proceeds measures are divided by initial debt and enter jointly.

Panel C restricts raw initial leverage to 5%–100% and 10%–40%, respectively. For a one-standard-deviation lower connected benchmark, holding initial leverage and the other controls fixed within the above-benchmark region, the positive-fraction effects are 2.47 and 3.35 percentage points (wild bootstrap $p = 0.079$ and $p = 0.194$). These restrictions exclude very low or unusually high initial leverage. The positive outcome combines the incidence and size of reductions. On the same 2,336 issuers in the 5%–100% range, retaining negative reductions for debt increases gives an effect of -3.08 points (95% confidence interval [-17.70, 11.55]; wild bootstrap $p = 0.696$).

A.10 Debt Adjustments with Extended Accounting Histories

Table A.10 extends the common-cohort comparison in Table 14 to issuers with observed debt at every quarter from -3 through $+4$. The 594 issuers form a selected accounting-history sample. For a one-standard-deviation lower connected benchmark among issuers starting above it, the IPO-quarter debt-change estimate is -9.36% of pre-IPO assets in raw debt

Table A.9: Initial Debt Reductions and Financing Intentions

Outcome	N	Effect (pp)	95% CR2 interval	CR2 p	Wild p
Panel A: Initial debt/assets $\geq 5\%$					
Any debt decline	2,602	-1.02	[-5.91, 3.87]	0.641	0.637
Fraction of initial debt reduced	2,602	2.46	[-0.29, 5.21]	0.073	0.067
Signed reduction fraction	2,602	-0.43	[-13.54, 12.67]	0.940	—
Signed fraction, winsorized	2,602	1.49	[-4.41, 7.39]	0.572	0.577
Explicit repayment plan	2,291	0.85	[-3.51, 5.21]	0.662	0.672
Primary-proceeds interaction	2,434	-0.20	[-0.64, 0.24]	0.263	0.340
Panel B: Initial debt/assets $\geq 10\%$					
Any debt decline	2,355	-1.05	[-5.54, 3.45]	0.604	—
Fraction of initial debt reduced	2,355	1.85	[-0.61, 4.31]	0.121	—
Signed reduction fraction	2,355	-2.09	[-16.01, 11.83]	0.737	—
Signed fraction, winsorized	2,355	0.82	[-3.81, 5.45]	0.692	—
Explicit repayment plan	2,057	1.00	[-3.86, 5.87]	0.644	—
Primary-proceeds interaction	2,218	-0.14	[-0.78, 0.51]	0.581	—
Panel C: Positive fractions under restricted initial leverage					
Initial leverage 5%–100%	2,336	2.47	[-0.35, 5.29]	0.078	0.079
Initial leverage 10%–40%	1,060	3.35	[-1.87, 8.56]	0.168	0.194

Notes: Reported effects correspond to a one-standard-deviation lower connected book leverage benchmark (11.80 percentage points, from Table 1) among issuers starting above the benchmark, holding issuer pre-IPO leverage and other controls fixed and keeping the pre-IPO gap positive. The scale is the SD of connected leverage, not of the positive gap. Samples require observed initial debt/assets at least 5% or 10%, without filling debt components. Balance outcomes require both components at both endpoints, positive initial assets and matching currencies. Listing is the first fiscal quarter-end on/after the offer within 270 days. The fraction reduced is $\max(D_{\text{pre}} - D_{\text{listing}}, 0)/D_{\text{pre}}$, retaining zeros. The signed fraction allows negative values for debt increases; the winsorized version clips it at its 1st/99th percentiles within the estimation sample. Binary outcomes are probability points; fractions are percentage points of initial debt. Repayment intentions use eligible pre-offer prospectuses; unavailable text is not a negative intention. The primary-proceeds rows report the change in the above-gap fraction association for an additional 100% of initial debt in primary gross proceeds, holding secondary proceeds fixed; both exposures are 1/99-winsorized. All regressions retain the starting-leverage spline, both connected gap components, conventional leverage gaps, issuer/connected/offering controls and SIC2/year/exchange fixed effects. Funding models interact each proceeds exposure with the connected gaps, initial-leverage spline and conventional gaps. CR2 intervals and p-values use SIC2/offer-year clustering and coefficient-specific Satterthwaite degrees of freedom. Wild tests use 4,999 requested null-imposed draws and CR1 two-way studentization; dashes indicate tests not computed. Panel C reports only positive fractions: the 5%–100% range includes both endpoints; the moderate range is $10\% \leq \text{PreLev} < 40\%$. These rows retain spline knots from the full IPO sample, and the moderate-range model also includes 10-point range fixed effects. The positive fraction combines the occurrence and size of reductions.

changes and -4.31% after winsorization (wild bootstrap $p = 0.031$ and $p = 0.034$). Net above-benchmark changes from quarter -3 through listing are not significant at 5% under either construction.

Table A.11 reports actual debt cash flows during the first year after the listing-quarter accounting date. A ten-percentage-point larger initial gap below connected leverage is associated with gross long-term debt issuance higher by 5.29% of pre-IPO assets and net long-term debt issuance higher by 3.16%, with wild cluster bootstrap p-values of 0.015 and 0.064, respectively. The net-issuance estimate is not significant at 5%, and its 95% confidence interval includes zero. The evidence is therefore stronger for gross borrowing activity than

Table A.10: Debt Adjustments in the Extended Accounting-History Sample

	Above connected leverage			Below connected leverage		
Period	Effect	95% interval	Wild p	Effect	95% interval	Wild p
Panel A: Raw debt changes						
Earlier quarter (−3 to −2)	+1.48	[-3.32, 6.28]	0.441	+1.22	[-3.13, 5.57]	0.472
Preceding quarter (−2 to −1)	+1.64	[0.18, 3.10]	0.132	+1.64	[0.35, 2.93]	0.033
IPO quarter (−1 to 0)	-9.36	[-17.49, -1.23]	0.031	+5.20	[-3.01, 13.41]	0.147
Post-listing (0 to +4)	-6.93	[-27.00, 13.15]	0.538	+17.46	[-5.20, 40.12]	0.064
IPO minus preceding quarter	-11.00	[-20.24, -1.76]	0.037	+3.56	[-5.62, 12.73]	0.353
Net change (−3 to 0)	-6.24	[-15.63, 3.16]	0.164	+8.06	[-1.46, 17.59]	0.043
Panel B: Winsorized debt changes						
Earlier quarter (−3 to −2)	+1.23	[-2.85, 5.30]	0.477	+0.50	[-3.38, 4.38]	0.727
Preceding quarter (−2 to −1)	+0.94	—	0.038	+1.24	[0.10, 2.37]	0.036
IPO quarter (−1 to 0)	-4.31	[-8.14, -0.48]	0.034	+1.83	[-0.47, 4.13]	0.073
Post-listing (0 to +4)	-6.42	[-25.32, 12.49]	0.543	+13.90	[-0.03, 27.82]	0.045
IPO minus preceding quarter	-5.69	[-10.47, -0.91]	0.040	+1.15	[-2.18, 4.48]	0.374
Net change (−3 to 0)	-1.05	[-7.00, 4.91]	0.658	+4.09	[-1.63, 9.81]	0.080

Notes: Effects are percentages of pre-IPO assets per one-standard-deviation change in connected book leverage (11.80 percentage points): a lower benchmark for issuers starting above it and a higher benchmark for those starting below it, holding initial leverage and other controls fixed. The sample contains the same 594 issuers in both panels. Periods and controls follow Table 14; the additional rows measure the earlier quarter and the net change through listing. Panel A uses raw debt changes; Panel B winsorizes each outcome at the 1st and 99th percentiles. Intervals use CR2/Satterthwaite inference clustered by SIC2 and offer year; wild-bootstrap p-values are two-sided. Dashes denote unavailable CR2 intervals. Appendices E and F provide construction and inference details.

for net debt accumulation. Both outcomes are winsorized at the 1st and 99th percentiles, use the same 3,295 issuers, and retain the initial-leverage and conventional-benchmark controls.

B IPO Pricing

Unless a table states otherwise, the supplementary results in this appendix use coefficient-specific Satterthwaite critical values with CR2 standard errors clustered by issuer SIC2 and offer month, and product-weight wild cluster bootstrap tests where reported. Because a few large industries dominate the cluster structure, the effective degrees of freedom for the baseline six-month connected return coefficient are 5.11. Consequently, a t-ratio of 2.98 in the baseline first-day regression corresponds to $p = 0.030$. Appendix F describes the bootstrap procedures and method comparisons.

Table B.1 examines weighting, same industry connections, and industry controls using the baseline six-month, top ten signal. Its baseline reproduces Table 4. Similarity weighting leaves the point estimate close to the baseline (0.283 compared with 0.282). Replacing separate industry and offer year fixed effects with SIC2-by-offer-year fixed effects yields a

Table A.11: Connected Leverage and First-Year Debt Issuance

	Gross LTD issuance (1)	Net LTD issuance (2)
Below connected leverage (BelowGap_C)	0.529**	0.316
	(0.184)	(0.165)
Wild bootstrap p-value	0.015	0.064
Above connected leverage (AboveGap_C)	-0.937**	-0.638***
	(0.271)	(0.172)
Wild bootstrap p-value	0.021	0.022
Effect of a 10-percentage-point larger below gap	5.29	3.16
95% confidence interval	[0.28, 10.30]	[-1.32, 7.64]
Flexible initial leverage and conventional benchmark gaps	Yes	Yes
Issuer, offering and connected-firm controls	Yes	Yes
SIC2, offer-year and exchange fixed effects	Yes	Yes
N	3,295	3,295

Notes: The outcomes are reported gross long-term debt issuance and net long-term debt issuance (issuance minus repayment), each divided by pre-IPO assets. The window begins at the listing-quarter accounting date and ends four fiscal quarters later; it does not begin on the exact listing day. Fiscal-YTD flows are de-cumulated over these endpoints with any required intervening year-end. Both columns use the same 3,295 issuers with observed issuance and repayment, valid fiscal-calendar alignment and matching reporting currency. Missing components are not imputed. Each outcome is winsorized separately at its sample's 1st and 99th percentiles after construction. Both connected gap sides enter jointly. Initial-leverage spline knots and basis are frozen on the 3,697-IPO financing cohort. Controls follow Table 13, Panel B, including industry and predicted-leverage gaps and issuer, offering and connected-firm characteristics. The economically scaled effects and intervals are percentages of pre-IPO assets per 10-percentage-point larger initial below-connected gap. CR2 standard errors in parentheses are two-way clustered by SIC2 and offer year; confidence intervals and stars use coefficient-specific Satterthwaite inference. Wild bootstrap p-values are two-sided and unadjusted for multiple testing, with 4,999 requested draws. Usable draws for the below-gap coefficient are 4,986 and 4,977 in columns (1) and (2), respectively; corresponding above-gap counts are 4,787 and 4,815. Net LTD issuance excludes net current-debt flows. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

connected return coefficient of 0.304, with a standard error of 0.101; the Satterthwaite and product-weight wild cluster bootstrap p-values are 0.041 and 0.046. Excluding same SIC2 comparables reduces the coefficient to 0.142 (standard error 0.060), significant at 10% under both procedures ($p = 0.066$ and $p = 0.078$, respectively). Offer price revision tests are reported in Section 4.

Table E.3 extends the aggregation comparison to the nine tests and within-family adjustments defined in Appendix F.

The market return coefficient conditions on correlated industry and connected returns, so its sign need not match the unconditional relation between market returns and IPO pricing.

B.1 Connected Set Size and Measurement Window

Table B.2 jointly varies the number of connected firms and the pre-offer measurement window. The first-day return coefficients remain positive and significant at 5% across the reported

Table A.12: Connected Financing News and First-Year Debt Cash Flows

	Gross LTD borrowing	LTD repayments	Net LTD borrowing
Panel A: Announcement proxy			
$N = 2,004$; issuers above the initial connected benchmark:	1,078		
Raw signal, controls, and cash flows	0.316	0.255	0.061
	(0.495)	—	(0.426)
p-value	0.528	—	0.887
Raw cash flows; winsorized signal and controls	-1.065	-2.062*	0.997
	(1.095)	(1.030)	(1.111)
p-value	0.339	0.055	0.377
Winsorized components; net as their difference	-0.849	-1.238**	0.388
	(0.936)	(0.471)	(1.196)
p-value	0.372	0.014	0.748
Each cash-flow outcome winsorized separately	-0.849	-1.238**	0.298
	(0.936)	(0.471)	(1.112)
p-value	0.372	0.014	0.791
Panel B: Periodic-filing dates			
$N = 2,059$; issuers above the initial connected benchmark:	1,114		
Raw signal, controls, and cash flows	-1.903	-2.220***	0.317
	(1.427)	(0.766)	(1.735)
p-value	0.193	0.007	0.856
Raw cash flows; winsorized signal and controls	-2.275	-1.961	-0.314
	(1.868)	(1.427)	(2.360)
p-value	0.233	0.180	0.895
Winsorized components; net as their difference	-0.327	-0.882	0.555
	(0.418)	(0.562)	(0.921)
p-value	0.441	0.127	0.551
Each cash-flow outcome winsorized separately	-0.327	-0.882	0.700
	(0.418)	(0.562)	(0.841)
p-value	0.441	0.127	0.412

Notes: Entries are financing-news slopes for issuers whose pre-IPO book leverage exceeds connected leverage at the start of the information window, estimated in pooled regressions with an above-benchmark interaction. The connected set and equal 1/10 weights are frozen at initial substantive registration. The information window begins at the later of registration and the saved last pre-IPO balance-sheet date and ends before the offer. Financing news aggregates quarterly debt changes over lagged assets less predictions based on public firms' lagged financing and fundamentals, estimated using only reports released in earlier calendar years. Panel A uses the earliest recorded earnings, news, or periodic-filing date; Panel B uses periodic-filing dates. These dates proxy item-level disclosure timing.

The outcomes cover the four quarters following the listing-quarter accounting date and share the pre-IPO asset denominator. Repayments are positive outflows. Net borrowing equals gross borrowing minus repayments and excludes current-debt flows. Within each panel, observations and regressors are identical across outcomes and transformations. The sample requires at least five usable original connected firms and observed outcomes and controls, without imputation; original weights are not renormalized. Controls include initial leverage splines, predetermined issuer characteristics, old connected and conventional leverage benchmarks, connected profitability and investment news, window length and disclosure exposure, and industry-by-offer-year fixed effects.

The first row uses raw signals, controls, and outcomes. Remaining rows use the winsorized prediction model and signal/control cutoffs from the prior repayment specification. Cash-flow winsorization uses common-sample 1st and 99th percentiles. The third row reconstructs net borrowing after winsorizing borrowing and repayment; the fourth winsorizes each outcome separately. The raw and reconstructed-net coefficients satisfy the borrowing-minus-repayment identity before rounding; separate net-outcome winsorization need not preserve it.

Parentheses report CR1 standard errors two-way clustered by SIC2 and offer year. Two-sided p-values use t critical values with the smaller cluster count minus one degrees of freedom and condition on the estimated news signal. Dashes denote nonpositive estimated variances. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table B.1: First-Day Returns: Alternative Specifications

	(1) Baseline	(2) Similarity weights	(3) Exclude issuer's SIC2 industry	(4) SIC2 × offer year FE
Connected return	0.282**	0.283**	0.142*	0.304**
	(0.095)	(0.094)	(0.060)	(0.101)
SIC2 industry return	0.183***	0.182***	0.342**	0.165**
	(0.044)	(0.044)	(0.093)	(0.057)
Market return	-0.610**	-0.610**	-0.636**	-0.580*
	(0.212)	(0.212)	(0.235)	(0.260)
Product-weight bootstrap p-value	0.050	0.049	0.078	0.046
IPO controls	Yes	Yes	Yes	Yes
Exchange FE	Yes	Yes	Yes	Yes
SIC2 FE	Yes	Yes	Yes	No
Offer year FE	Yes	Yes	Yes	No
SIC2 × offer year FE	No	No	No	Yes
N	3,954	3,954	3,875	3,954
Adjusted R^2	0.275	0.275	0.261	0.222

Notes: The dependent variable is the first-day closing price divided by the offer price minus one. The connected return is the equal-weighted return of the ten connected firms from six calendar months before the recorded offer date through the market close immediately before the first public trading day. Column (1) repeats Table 4; columns (2)–(4) use cosine similarity weights, exclude connected firms in the issuer's SIC2 industry, and replace separate SIC2 and offer year fixed effects with SIC2-by-offer-year fixed effects, respectively. All regressions include SIC2 industry and market returns, the IPO controls of Section 4.1, and the fixed effects shown. Continuous variables are winsorized at the 1st and 99th percentiles. Parentheses contain two-way CR2 standard errors clustered by issuer SIC2 and offer month; stars use coefficient-specific Satterthwaite tests. The displayed bootstrap p-values use products of independent SIC2 and offer-month Rademacher weights. Appendix F provides the method comparison. Table E.3 compares aggregation weights for the nine tests defined there. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

specifications.

Table B.3 repeats the bookbuilding revision regressions of Section 4 with five to thirty connected firms.

B.2 Bookbuilding Timing and Alternative Comparisons

The baseline bookbuilding return ends at $T-2$, excluding the final session before first public trading. Table B.4 compares this endpoint with $T-1$ and $T-3$, using both available complete cases and the same 3,824 IPOs across endpoints. With each return cutoff, issuers with higher connected returns have higher offer price revisions and first-day returns. The first-day result remains after controlling for the revision, and all corresponding connected return coefficients are significant at 5% under both inference procedures. The earlier endpoint reduces exposure to returns realized after pricing.

Table B.2: Connected Returns by Number of Connected Firms and Pre-IPO Window

	$K = 5$	$K = 10$	$K = 15$	$K = 20$	$K = 25$	$K = 30$
Panel A: First-day return						
$L = 1$	0.532***	0.646**	0.892**	0.992**	0.970***	1.051***
	(0.096)	(0.197)	(0.290)	(0.298)	(0.263)	(0.266)
$L = 3$	0.339***	0.454**	0.555**	0.608**	0.595**	0.634**
	(0.084)	(0.134)	(0.178)	(0.174)	(0.167)	(0.177)
$L = 6$	0.205**	0.282**	0.376**	0.400**	0.402**	0.434***
	(0.057)	(0.095)	(0.127)	(0.124)	(0.111)	(0.112)
$L = 12$	0.081***	0.117**	0.158**	0.178**	0.172**	0.189**
	(0.019)	(0.035)	(0.054)	(0.058)	(0.050)	(0.059)
Panel B: 12-month BHAR						
$L = 1$	-0.047	-0.176	-0.082	-0.064	-0.166	-0.298
	(0.102)	(0.125)	(0.156)	(0.176)	(0.176)	(0.205)
$L = 3$	0.038	-0.073	-0.073	-0.097	-0.107	-0.137
	(0.077)	(0.124)	(0.135)	(0.148)	(0.178)	(0.192)
$L = 6$	-0.031	-0.075*	-0.063	-0.096	-0.110	-0.129
	(0.030)	(0.037)	(0.048)	(0.059)	(0.078)	(0.095)
$L = 12$	-0.029**	-0.079***	-0.071**	-0.108**	-0.112**	-0.122*
	(0.011)	(0.018)	(0.026)	(0.037)	(0.044)	(0.053)
Panel C: 24-month BHAR						
$L = 1$	-0.068	-0.340	-0.300	-0.315	-0.374	-0.581
	(0.173)	(0.201)	(0.240)	(0.293)	(0.323)	(0.348)
$L = 3$	0.002	-0.186	-0.165	-0.176	-0.108	-0.173
	(0.092)	(0.109)	(0.106)	(0.123)	(0.148)	(0.159)
$L = 6$	-0.090	-0.110	-0.105	-0.141	-0.100	-0.126
	(0.046)	(0.069)	(0.067)	(0.074)	(0.083)	(0.090)
$L = 12$	-0.069**	-0.106**	-0.113**	-0.152**	-0.122*	-0.133*
	(0.020)	(0.037)	(0.040)	(0.047)	(0.058)	(0.065)
Panel D: 36-month BHAR						
$L = 1$	-0.160	-0.429*	-0.317	-0.387	-0.517	-0.744**
	(0.170)	(0.179)	(0.228)	(0.301)	(0.350)	(0.280)
$L = 3$	0.101	-0.127	-0.118	-0.103	-0.115	-0.140
	(0.124)	(0.083)	(0.115)	(0.122)	(0.154)	(0.165)
$L = 6$	-0.031	-0.070	-0.107	-0.117	-0.102	-0.114
	(0.052)	(0.043)	(0.058)	(0.080)	(0.089)	(0.098)
$L = 12$	-0.045	-0.088**	-0.108*	-0.141*	-0.113	-0.125
	(0.025)	(0.032)	(0.043)	(0.059)	(0.062)	(0.065)

Notes: Each cell reports the coefficient on the return averaged over K connected firms from L calendar months before the recorded offer date through the market close immediately before the first public trading day. Panel A uses the first-day return; Panels B–D use issuer market-adjusted BHAR from the first-day close. All regressions include SIC2 industry and market returns over the same dates and exchange, SIC2, and offer year fixed effects. IPO controls follow Section 4.1 for Panel A and Appendix E for Panels B–D. Return construction follows Appendix E. The baseline uses $K = 10$ and $L = 6$; N ranges from 3,821 to 3,954. Continuous variables are winsorized at the 1st and 99th percentiles. Standard errors in parentheses are CR2, two-way clustered by issuer SIC2 and offer month. Stars use coefficient-specific Satterthwaite inference: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table B.3: Offer Price Revisions During Bookbuilding by Connected Set Size

	K = 5	K = 10	K = 15	K = 20	K = 30
Connected return, pre-range	0.049***	0.066***	0.090***	0.093***	0.087***
	(0.003)	(0.013)	(0.016)	(0.016)	(0.021)
Connected return, bookbuilding	0.065**	0.133***	0.178**	0.246***	0.304***
	(0.020)	(0.035)	(0.059)	(0.048)	(0.061)
Observations	3,830	3,830	3,830	3,830	3,830

Notes: The dependent variable is the log offer price relative to the midpoint of the initial filing range. Connected returns average K firms over the pre-range and bookbuilding windows defined in Table 7; the sample requires an observed range date. The $K = 10$ column reproduces column (1) of that table. Regressions include market and SIC2 industry returns over both windows, the log of one plus the number of bookbuilding trading days, the IPO controls of Section 4.1, and exchange, SIC2, and offer year fixed effects. Continuous variables are winsorized at the 1st and 99th percentiles. Standard errors in parentheses are CR2, two-way clustered by issuer SIC2 and offer month. Stars use coefficient-specific Satterthwaite inference: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table B.4: Bookbuilding Return Cutoff Sensitivity

Cutoff	IPOs	Revision	First-day	First-day given revision
Available complete cases				
$T-1$	3,838	0.117** (0.036)	0.363** (0.113)	0.303** (0.095)
		[0.019]	[0.030]	[0.035]
$T-2$	3,830	0.133*** (0.035)	0.367** (0.123)	0.299** (0.103)
		[0.011]	[0.036]	[0.043]
$T-3$	3,824	0.157*** (0.037)	0.388** (0.125)	0.307** (0.104)
		[0.008]	[0.030]	[0.037]
Common across cutoffs				
$T-1$	3,824	0.116** (0.036)	0.363** (0.113)	0.303** (0.095)
		[0.023]	[0.032]	[0.034]
$T-2$	3,824	0.133*** (0.035)	0.368** (0.123)	0.299** (0.104)
		[0.012]	[0.039]	[0.042]
$T-3$	3,824	0.157*** (0.037)	0.388** (0.125)	0.307** (0.104)
		[0.008]	[0.030]	[0.037]

Notes: Each cell reports the connected bookbuilding return coefficient, CR2 standard error in parentheses, and two-way wild cluster bootstrap p-value in brackets. All portfolios, including market and industry controls, share the stated session endpoint relative to first public trading. The top panel uses available complete cases; the lower panel holds IPOs fixed across endpoints. All other regressors and transformations follow Table 7, including full range-to-trading duration. First-day returns omit the revision in the middle column and include it in the final column. CR2/Satterthwaite and wild inference cluster by issuer SIC2 and offer month. Stars follow the Satterthwaite tests. No missing returns are imputed.

Table 8 compares connected returns with industry, size-matched, and characteristic-matched returns using the pre-range and bookbuilding windows. Conditional on the offer price revision, the connected bookbuilding coefficients are 0.285, 0.293, and 0.304. Their differences from the alternative coefficients are significant at 5% under both inference procedures, with wild cluster bootstrap p-values of 0.042, 0.029, and 0.023. Connected returns remain significant at 5% in the revision regressions alongside size- and characteristic-matched returns, but not at 10% alongside SIC4 returns. The revision comparisons do not reject equal coefficients at 5%. Same-industry bookbuilding returns predict revisions; the SIC4 alternative also has a positive conditional first-day coefficient. The SIC4, size-matched, and characteristic-matched comparisons use 3,324, 3,812, and 3,426 of the 3,830 IPOs, respectively.

Table B.5 compares connected returns with returns on public firms explicitly identified as competitors in the issuer's pre-IPO disclosures. In joint regressions for 2,292 IPOs, one-standard-deviation increases in named-competitor and connected returns during bookbuilding are associated with offer prices approximately 2.71% and 0.51% higher relative to the initial range midpoint, respectively ($p = 0.017$ for the difference). The unconditional first-day effects do not differ significantly, and a direct comparison of the named-firm advantage across the two pricing stages does not reject equality ($p = 0.288$). On a common sample of 1,948 IPOs, restricting named competitors to the eligibility universe used for text matching reduces the difference between the approximate percentage price responses from 1.66 to 0.21 percentage points; the latter is statistically insignificant ($p = 0.782$).

Table B.6 reports portfolio and pricing-stage comparisons for the named-competitor evidence in Section 4.3. The common-universe comparison holds the 1,948-IPO sample fixed while changing named-competitor eligibility. The disjoint comparison removes overlap from the two original portfolios. Cross-stage and revision-share tests retain the joint covariance across signals and pricing outcomes; neither rejects equality at 5%.

B.3 Completion Selection

Table B.7 compares pricing coefficients across thirds of predicted completion probability. Five-fold logit predictions keep every issuer's attempts in the same fold and use pre-range connected, market and industry returns, the initial range midpoint and entry-year fixed effects. The thirds are formed over the sample of completed and withdrawn IPOs. The table

Table B.5: Named Competitors and Text-Based Connected Firms in IPO Pricing

	(1) Offer price revision	(2) First-day return
Named-competitor return, bookbuilding	2.71	3.86
p-value	0.001	0.002
Connected return, bookbuilding	0.51	3.19
p-value	0.253	0.068
Difference: named − connected	2.20	0.67
p-value	0.017	0.573
Named-competitor return SD (pp)	11.67	11.67
Connected return SD (pp)	11.47	11.47
Pre-range returns and other controls	Yes	Yes
Industry, offer-year, and exchange effects	Yes	Yes
Observations	2,292	2,292

Notes: The unit of observation is an IPO. Column (1) uses the log offer price relative to the midpoint of the initial filing range; column (2) uses the log first-day close relative to the offer price and does not condition on the offer price revision. Both regressions jointly include returns on all recovered named public competitors and the ten closest text-based connected firms during bookbuilding. The portfolios may overlap: named competitors need not belong to the connected set, and the connected set retains its named members. Membership is fixed using disclosures available by the initial range filing. Bookbuilding returns run from that filing through the $T-2$ close; pre-range returns cover the preceding six months. Portfolios equally weight available individual cumulative simple returns, requiring observations on at least 60% of expected trading days for each constituent. Missing returns are not imputed. The common sample requires both portfolios' returns over both windows, outcomes, and controls. The recovered named set need not be exhaustive. Other regressors are both portfolios' pre-range returns, market and SIC2 industry returns over both windows, log one plus bookbuilding trading days, log initial range midpoint, log one plus issuer age, venture or growth-capital backing, and an Internet or high-technology indicator, together with SIC2, offer-year, and exchange fixed effects. The industry portfolios exclude the issuer's thirty closest connected firms. Continuous variables retain the existing 1st and 99th percentile winsorization. Reported effects approximate percentage changes in the corresponding price ratios per one-standard-deviation increase in each winsorized return, calculated as 100 times the coefficient times its sample SD across these 2,292 IPOs. The difference row reports $100(s_N\widehat{\beta}_N - s_C\widehat{\beta}_C)$, an approximate percentage-point difference between the price responses, and tests equality of the two one-SD effects using the full joint covariance, treating the observed SDs as fixed. All *p*-values are two-sided and use CR2 standard errors clustered by issuer SIC2 and offer month with Satterthwaite degrees of freedom.

reproduces the original pricing models, then reports the linked sample and fully interacted probability groups. The completion model uses issuer-level inputs available by the original range filing. These retrospective comparisons condition on the estimated groups and do not account for uncertainty in their construction. They assess sensitivity among completed offerings.

For the completion diagnostic, I form connected sets for both completed and withdrawn attempts at initial registration, applying the same historical identity, active listing, and preceding-session trade requirements. Each attempt retains this ranking for later return windows. The base withdrawal cohort contains 1,324 attempts; availability of a dated range

Table B.6: Explicit Competitor Recognition: Portfolio and Pricing-Stage Comparisons

Comparison	IPOs	Named	Connected	Difference	p-value
Panel A: Offer-price revision effects					
Baseline	2292	2.71	0.51	2.20	0.017
Common sample, original sets	1948	2.76	1.10	1.66	0.085
Common eligibility universe	1948	1.64	1.42	0.21	0.782
Disjoint portfolios	2160	2.96	0.48	2.48	0.019
Panel B: Pricing-stage and share comparisons					
Contrast			Estimate	95% interval	p-value
Named advantage: revision less first-day			1.53	[-1.77, 4.82]	0.288
Revision share: named minus connected (pp)			27.42	[-199.1, 218.6]	0.397

Notes: Panel A reports approximate percentage changes in the offer price relative to the initial range midpoint per one sample SD of each winsorized bookbuilding return, calculated as 100 times the coefficient times the SD. Differences between responses are in approximate percentage points. All comparisons jointly include both portfolios and their pre-range returns, with the controls, fixed effects, return construction and $T-2$ endpoint of Table B.5. The full connected top ten retains named members. The common eligibility sample contains 1,948 IPOs; its original-portfolio row holds those IPOs fixed before restricting named competitors to firms eligible for text matching as of the range filing. The disjoint comparison removes overlap from both portfolios. Two-sided coefficient tests use SIC2/offer-month CR2 with Satterthwaite inference. Panel B uses all 2,292 baseline IPOs. The cross-stage contrast is the named-minus-connected revision effect less the corresponding unconditional first-day effect; its interval uses the full joint covariance and Satterthwaite reference. Revision shares divide each revision coefficient by its revision-plus-unconditional-first-day coefficient. The share difference is in percentage points, with a basic 95% interval and two-sided p-value from 2,000 paired SIC2/offer-month pigeonhole bootstrap draws. All ratio draws are retained, including weak or negative denominators.

or other model inputs determines the reported estimation sample, without changing that cohort. The completion predictions use 4,071 attempts, including 557 withdrawals, and the linked bookbuilding comparison contains 3,274 completed IPOs. The pricing regressions retain their initial-range connected sets.

B.4 Business Similarity and IPO Price Formation

Table B.8 examines whether the relation between connected stock returns and IPO pricing varies with business similarity. I partition the thirty highest-ranked connected firms into three nonoverlapping groups, comprising ranks 1–10, 11–20, and 21–30. The groups use the similarity ranking fixed at the range filing. Their returns during bookbuilding enter jointly, together with the corresponding pre-range returns and the controls in Table 7. All specifications use the same 3,828 IPOs.

No pairwise coefficient difference is significant at 5% for any reported pricing outcome. For total price updates, the closest-minus-middle and closest-minus-distant tests have p-values of

Table B.7: IPO Pricing by Predicted Completion Probability

	Original	Linked	Low	Middle	High	Equal slopes p
Panel A: Bookbuilding						
Offer-price revision	1.332***	1.297**	1.166	1.836	0.988	1.000
	(0.346)	(0.447)	(0.919)	–	(0.731)	
First-day log return	3.675**	3.423**	3.581*	1.918	5.076**	0.415
	(1.229)	(1.164)	(1.519)	(1.036)	(1.762)	
First-day log return, conditional	2.990**	2.746**	2.979*	1.031	4.441**	0.261
	(1.035)	(0.862)	(1.349)	(0.946)	(1.467)	
IPOs	3,830	3,274	898	1,139	1,237	
Panel B: Six-month pre-offer returns						
Simple first-day return	2.817**	2.738**	5.288*	2.907*	1.250	0.322
	(0.946)	(0.784)	(2.202)	(1.174)	(0.672)	
Offer-price revision	1.052**	1.162***	1.479*	1.171**	0.901***	0.743
	(0.285)	(0.285)	(0.671)	(0.264)	(0.216)	
First-day log return	1.592**	1.574**	2.978*	1.656*	0.825*	0.328
	(0.492)	(0.409)	(1.205)	(0.667)	(0.373)	
IPOs	3,954	3,275	899	1,139	1,237	
Panel C: Bookbuilding revision asymmetry, sign-intercept model						
Positive-return slope	0.512	0.738	0.800	1.600*	−0.561	
	(0.494)	(0.520)	(0.913)	(0.656)	(0.694)	
Negative-return slope	4.249***	4.218**	2.932	6.800***	7.839**	
	(0.856)	(1.574)	(2.621)	(0.771)	(2.404)	
Negative minus positive	3.737***	3.480**	2.133	5.200***	8.399***	0.290
	(0.777)	(1.237)	(3.050)	(1.087)	(2.125)	

Notes: Entries are outcome changes for a 10-percentage-point increase in connected returns. The simple first-day return is in percentage points; other entries approximate percentage changes in the corresponding price ratios, calculated as 100 times the log-outcome slope times the return change. Contrasts are differences between these approximate responses. Bookbuilding ends at $T-2$; six-month pre-offer windows and all winsorized variables retain the original definitions. Each row reproduces the original specification, restricts it to linked observations, and estimates it within probability thirds. Completion scores use five-fold issuer-grouped logit predictions from pre-range connected, market and SIC2 returns, log initial midpoint and entry-year effects. Thirds are fixed over 4,071 retained attempts. Prediction information must be available by the original range date. The scores are retrospective and conditional on the sample of completed and withdrawn IPOs. Original controls and exchange, SIC2 and offer-year effects are included separately in each stratum. Panel C additionally allows a negative-return intercept. Standard errors in parentheses are SIC2/offer-month CR2; dashes denote unavailable CR2 inference, with point estimates and observations retained. Stars use contrast-specific Satterthwaite tests. The final column is a conservative Bonferroni test of two cross-group contrasts spanning equality, including shared-cluster covariance. In Panel C it tests equality of the asymmetry. Inference conditions on fitted groups and omits score-estimation uncertainty. *** $p<0.01$, ** $p<0.05$, * $p<0.10$.

0.930 and 0.856.

Table B.8: Connected Returns During Bookbuilding by Business Similarity Rank

	(1) Revision	(2) First-day	(3) Total update	(4) First-day conditional
Ranks 1–10	0.033	0.257	0.290	0.241*
	(0.049)	(0.134)	(0.173)	(0.116)
Ranks 11–20	0.113*	0.199**	0.312**	0.141*
	(0.055)	(0.077)	(0.114)	(0.070)
Ranks 21–30	0.156***	0.100	0.256***	0.020
	(0.042)	(0.053)	(0.060)	(0.062)
Pre-range returns and IPO controls	Yes	Yes	Yes	Yes
Market and industry returns	Yes	Yes	Yes	Yes
Fixed effects	Yes	Yes	Yes	Yes
Revision control	No	No	No	Yes
Observations	3,828	3,828	3,828	3,828
Adjusted R^2	0.266	0.270	0.323	0.410
Coefficient differences: two-sided p-values in brackets				
Ranks 1–10 minus 11–20	-0.080 [0.403]	0.058 [0.753]	-0.022 [0.930]	0.099 [0.565]
Ranks 1–10 minus 21–30	-0.124 [0.170]	0.157 [0.275]	0.034 [0.856]	0.221 [0.139]
Ranks 11–20 minus 21–30	-0.044 [0.527]	0.099 [0.405]	0.055 [0.721]	0.121 [0.279]

Notes: The unit of observation is an IPO. Connected firms are partitioned into nonoverlapping groups using text-based similarity ranks fixed at the initial range filing. Each group return equally weights constituent buy-and-hold returns. Bookbuilding runs from the initial range filing through the close two trading sessions before first public trading ($T-2$); the pre-range window covers the preceding six calendar months. Within each window, constituent returns require coverage of at least 60% of trading days; coverage screens do not reassign ranks or replace firms. All three bookbuilding returns and their corresponding pre-range returns enter jointly. Outcomes are log offer price relative to the initial range midpoint (Revision), log first-day close relative to offer price (First-day), and their sum (Total update). Column (4) additionally controls for Revision. All columns retain the market and SIC2 industry returns over both windows, bookbuilding duration, IPO controls, and SIC2, offer-year, and exchange fixed effects of Table 7. The same 3,828 IPOs enter every column. Continuous variables retain the original winsorization; group returns are winsorized at the 1st and 99th percentiles of the full construction panel before the common-sample restriction. Revision and first-day outcomes retain separate winsorization, and Total update is their sum. Parentheses report CR2 standard errors, two-way clustered by issuer SIC2 and offer month. The lower panel reports direct coefficient differences using the joint covariance matrix, with two-sided p-values in brackets. Stars use coefficient-specific Satterthwaite inference: *** $p<0.01$, ** $p<0.05$, * $p<0.10$.

B.5 Filing Uncertainty and Connected Announcement News

Filing uncertainty. I link the issuer's selected registration accession to its CIK and SEC file number, then select the earliest observed original S-1 or F-1 for that registration episode. The filing must be dated no later than the initial range filing. I require a cached full document with matching source metadata and retain its accession, filing date, URL, and source hash.

Amendments and final prospectuses are not substituted for an unavailable initial filing. Since eligible announcement windows begin strictly after the initial range date, the filing language precedes the measured news.

I extract visible text from the primary registration document, excluding submission exhibits, scripts, styles, comments, and explicitly hidden HTML. When a Part II heading is recognized, the text ends before the last such heading, avoiding its earlier table-of-contents entry. A boundary is recognized for 3,476 of the 3,632 retained filings; the remaining 156 use the full visible primary-document text and retain a boundary-not-found flag. Uncertainty is the number of token occurrences in the active uncertainty category of the Loughran–McDonald Master Dictionary, 1993–2025 version, divided by all alphabetic tokens matching `[A-Za-z]+`. The cached dictionary is the authors' March 2026 release and is used retrospectively. Capitalized `May` is excluded from the uncertainty numerator. The measure is not trimmed or winsorized. I subtract its mean, 0.013176, and divide by its sample standard deviation, 0.002321, calculated from the common 3,632-IPO estimation sample to obtain Z_i.

Announcement news. Earnings-announcement dates are Compustat quarterly report dates (RDQ), linked to CRSP securities using dated CCM links of types LC, LU, and LS with primary-link designations P or C. Conflicting dates and ambiguous firm links remain flagged rather than being resolved by choosing a record; only unambiguous linked dates enter the measured windows. Reported RDQs preceding the fiscal-quarter end are retained with a timing flag. An announcement dated on a nontrading day is assigned to the next market session. Each event window comprises that session and the immediately preceding and following sessions, $[-1, +1]$. Every window must begin strictly after the initial range date, end by the $T-2$ close before first public trading, and finish strictly before the recorded offer date.

The daily adjusted return is the residual from a model of the connected stock's excess return on the Fama–French five factors, momentum, and its leave-security-out SIC2 industry excess return, with an intercept. The industry return equally weights finite stock returns using SIC2 observed in the preceding calendar month. Each model uses up to 252 market sessions strictly before the initial range date, requires at least 120 complete observations and a full-rank design, and retains these coefficients during bookbuilding. The announcement

signal is

$$S_i^{EA} = \sum_{j \in C_i} w_{ij} \sum_{t \in BB_i} \widetilde{r}_{jt} \, \mathbf{1}\{t \text{ belongs to an eligible earnings window for } j\}.$$

The connected ranking and original weights, $w_{ij} = 0.1$, are fixed at the initial range. Overlapping windows for the same stock contribute each return date once. I do not replace firms or renormalize weights to the announcing firms or those with measured returns.

I sum observed contributions at these fixed weights. A firm represented in the dated announcement registry with no eligible event contributes zero; a firm without valid dated-registry coverage has an unavailable announcement contribution. An event requires all three adjusted returns to contribute a measured announcement return. Missing contributions are left unmeasured while other observed contributions remain in the issuer-level sum. An entirely unmeasured signal remains missing. Registry presence does not establish complete historical announcement coverage. IPOs with no observed eligible announcement remain in the regressions when their required variables are observed. The nonannouncement signal sums observed adjusted returns outside the union of eligible event-window dates, including returns associated with announcements whose windows cross a boundary. Neither return signal is additionally trimmed or winsorized.

Announcement exposure is measured by the original weight assigned to firms with an eligible event and by the eligible event count multiplied by 0.1. Coverage controls include the original weight represented in the dated registry; the weight with an observed RDQ on or before the range and on or after the exclusive measurement endpoint; the fixed-weight fraction of linked quarterly source rows with missing RDQ; the fixed-weight fraction of expected post-range return sessions observed; the complete-return event count multiplied by 0.1; and the total selected original portfolio weight. Source-row missingness describes the linked history, rather than a verified count of missing announcements within bookbuilding. Empty return windows retain their arithmetic zero and are distinguished from unavailable measurements in nonempty windows.

Estimation sample. Table B.9, Panel A, reports sequential requirements relative to the 3,830 IPOs in the dated bookbuilding sample. 99 IPOs have registration forms other than S-1/F-1, and an original S-1/F-1 cannot be recovered for another 99. The remaining 3,632 IPOs have all required model variables. They include 1,431 IPOs with no eligible connected announcement. Panel B estimates the announcement-news baseline on all 3,830 eligible IPOs

and on these exact 3,632 observations. Both baseline specifications omit filing uncertainty and every interaction involving uncertainty, while retaining the other regressors and their common composition reference.

Specification and inference. In addition to the terms displayed in equation (6), the model includes connected nonannouncement returns and their interaction with Z_i; connected returns over the six months preceding the range; market and SIC2 industry returns before the range and during bookbuilding; and interactions of all four market and industry return measures with Z_i. Issuer controls are the log initial range midpoint, log one plus age, venture-capital or growth-capital backing, an Internet or high-technology indicator, and log one plus bookbuilding trading days. These existing controls retain their original winsorization. The model also includes the announcement-exposure and coverage measures above, exchange, SIC2, and offer-year fixed effects, and predetermined measures of connected-firm composition: the fixed-weight exposure to the issuer's SIC2 industry, the weight with observed SIC2, and the fixed-weight sum of text similarities. Each composition measure also interacts with announcement news. I center these measures at their common 3,632-IPO sample means, so the announcement-news coefficient is evaluated at that mean composition.

The outcomes are the literal log-price changes in equation (2), without additional winsorization. The first-day-return specification does not condition on the offer price revision. All three regressions use identical observations, regressors, and transformations, and total repricing is constructed as the sum of the two pricing components. Standard errors use the two-way CR2 covariance estimator clustered by issuer SIC2 and offer month. Tests are two-sided and condition on the constructed return signals and text scores. Table 9's total-repricing interaction uses the null-imposed SIC2-weight bootstrap ($p = 0.0040$; Appendix F). Other tests and Table B.9 intervals use CR2/Satterthwaite inference; the analytic interaction test has 6.083 degrees of freedom.

Marginal responses. Panel C reports the response to a one-standard-deviation increase in announcement news at the 25th, 50th, and 75th percentiles of standardized filing uncertainty. Holding industry and similarity composition at their common-sample means, the response for outcome Y is

$$100 s_{EA}\big(\widehat{\beta}_Y + \widehat{\gamma}_Y z_q\big),$$

Table B.9: Uncertainty Analysis: Sample Construction and Announcement-News Responses

Panel A: Sample construction

Sample requirement	Excluded	Remaining
Dated bookbuilding sample	—	3,830
S-1/F-1 registration scope	99	3,731
Verified original filing and usable text	99	3,632
Total excluded	198	

Panel B: Announcement-news baseline across samples

	Largest eligible announcement sample	Retained uncertainty sample
Offer-price revision		
Connected announcement news	0.109	0.246
	(0.456)	(0.455)
First-day return		
Connected announcement news	0.007	0.022
	(0.187)	(0.184)
Total repricing		
Connected announcement news	0.116	0.268
	(0.578)	(0.557)
Observations	3,830	3,632

Panel C: Response to one-SD announcement news by uncertainty

Uncertainty percentile	z_q	Offer-price revision	First-day return	Total repricing
25th	-0.688	-0.13	-0.32	-0.46
95% confidence interval		[-1.65, 1.38]	[-1.40, 0.75]	[-2.62, 1.71]
50th	-0.022	0.77	0.28	1.05
95% confidence interval		[-1.21, 2.74]	[-0.66, 1.23]	[-1.33, 3.43]
75th	0.637	1.66	0.89	2.55
95% confidence interval		[-1.10, 4.42]	[-0.41, 2.18]	[-0.58, 5.67]

Notes: Panel A applies requirements sequentially to the 3,830-IPO dated bookbuilding sample; each excluded IPO is counted once. Panel B estimates otherwise identical announcement-news baselines on the two samples. Both omit filing uncertainty and every interaction involving uncertainty, while retaining nonannouncement returns, the other controls and fixed effects, and the same industry/similarity composition reference used in Table 9. Observed no-announcement IPOs remain included. Parentheses report standard errors.

Panel C uses the Table 9 model and reports $100 s_{EA}(\widehat{\beta}_Y + \widehat{\gamma}_Y z_q)$ as approximate percentage changes in the corresponding price ratios, holding industry and similarity composition at their common-sample means. Both z_q and the single news SD, $s_{EA} = 0.019478$, are calculated from the common 3,632-IPO sample. Confidence intervals use the joint covariance of the announcement-news and uncertainty-interaction coefficients. CR2 covariance estimates are two-way clustered by issuer SIC2 and offer month; intervals use contrast-specific Satterthwaite critical values. Panel B stars use the corresponding coefficient-specific two-sided tests. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

where z_q is the relevant pooled percentile and $s_{EA} = 0.019478$ is the announcement-news standard deviation in the same 3,632-IPO sample. I use this news scale for every outcome and uncertainty percentile. Confidence intervals use the joint CR2 covariance of the announcement-news coefficient and its uncertainty interaction, including their covariance, and contrast-specific Satterthwaite critical values.

C Return Comovement

C.1 Return Construction and Similarity Ranks

Table C.1 re-estimates the baseline comovement specification of Table 15, column (1). Dropping connected firms from the portfolio after their delisting month, rather than assigning them the market return, yields a connected return coefficient of 0.275 (standard error 0.017). Winsorizing issuer, connected firm, and industry monthly returns at the 1st and 99th percentiles yields a coefficient of 0.337 (standard error 0.022). Restricting the baseline to months 1–12 or 13–36 gives coefficients of 0.294 (0.029) and 0.339 (0.028), respectively. The connected return adds explanatory power beyond the industry return under both constructions. In each column, issuers remain in the panel through the delisting month, including the delisting return, and leave thereafter. Standard errors are two-way clustered by issuer and calendar month.

Finer similarity ranks. Partitioning connected firms into ranks 1–5, 6–10, 11–20, and 21–30 yields joint return coefficients of 0.174, 0.099, 0.145, and 0.114, respectively, all significant at 1%. The comparison uses the same 3,939 issuers and 127,864 issuer-months as the three-band equal-weighted specification in Table 16. The coefficients do not decline monotonically with similarity: ranks 11–20 have a larger coefficient than ranks 6–10. Unequal band widths also allow portfolio diversification to affect coefficient comparisons.

C.2 Issuer-Disclosed Firms and Portfolio Diagnostics

Table C.2 partitions the connected set C_i and recovered issuer-disclosed set D_i into connected-only firms ($C_i \setminus D_i$), named-only firms ($D_i \setminus C_i$), and overlap firms ($C_i \cap D_i$). Their joint coefficients are 0.195, 0.212, and 0.144, respectively, each significant at the 1% level. The common sample requires all three returns and contains 809 issuers and 26,020 issuer-months,

Table C.1: Post-Listing Return Comovement: Alternative Constructions

	(1) Baseline	(2) Surviving connected firms	(3) Winsorized returns
Connected-firm return	0.330***	0.275***	0.337***
	(0.023)	(0.017)	(0.022)
SIC2 industry return	0.739***	0.728***	0.682***
	(0.027)	(0.026)	(0.024)
Issuer FE	Yes	Yes	Yes
Calendar-month FE	Yes	Yes	Yes
R^2 net of fixed effects	0.033	0.034	0.041
R^2 net of fixed effects, industry only	0.026	0.026	0.031
Incremental R^2 from the connected return	0.007	0.008	0.010
Issuer-months	127,963	127,946	127,963
Issuers	3,942	3,942	3,942

Notes: Each column re-estimates the baseline specification of Table 15, column (1), over event months 1 through 36. Column (1) repeats that estimate. Column (2) drops a connected firm from the portfolio after its delisting month instead of assigning it the market return, so the connected return averages over surviving firms; issuer-months whose connected firms have all delisted leave the sample. Column (3) winsorizes the issuer, connected-firm, and industry monthly returns at the 1st and 99th percentiles of the column (1) estimation sample. The industry-only R^2 and the incremental R^2 are computed on each column's own sample. Standard errors in parentheses are two-way clustered by issuer and calendar month. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

a selected subset of the 2,358-issuer disclosure comparison.

Table C.2: Common-Sample Decomposition of Post-IPO Comovement

	(1)
Connected-only return	0.195***
	(0.032)
Named-only return	0.212***
	(0.028)
Overlap return	0.144***
	(0.015)
SIC2 return	Yes
Issuer FE	Yes
Calendar-month FE	Yes
R^2 net of fixed effects	0.049
Issuers	809
Issuer-months	26,020

Notes: The three disjoint portfolio returns enter jointly with SIC2 returns and issuer and calendar-month fixed effects. Connected-only firms are absent from recovered issuer-named competitor labels; named-only firms are outside the frozen connected set; overlap firms belong to both. Returns cover months 1–36 after listing and retain the construction of Table 15. All three returns must be observed on the same issuer-month. The resulting 809 issuers and 26,020 observations form a selected subset of the 2,358 issuers in that table, column (2). Table C.3 reports portfolio sizes and standardized effects. R^2 removes both fixed effects. Standard errors in parentheses are two-way clustered by issuer and calendar month. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table C.3 summarizes portfolio sizes and standardized return effects. Overlap is absent for 60.2% of the 2,358 issuers in the broad comparison; requiring all three portfolios reduces the sample to 809 issuers. On the same 26,020 issuer-months as Table C.2, one-standard-deviation increases in connected-only, named-only, and overlap returns correspond to 2.30, 2.59, and 2.53 percentage points of issuer return. These SDs describe the smaller decomposition sample; the one-SD magnitudes in Section 6 use the corresponding regression samples. Standardization leaves fitted values and inference unchanged. It puts the effects on a common SD scale.

Table C.3: Portfolio Sizes and Standardized Comovement

Panel A: Portfolio sizes at formation

Portfolio	Mean	Median	Zero (%)
Broad comparison: 2,358 issuers			
Connected-only	9.28	10	0.0
Named-only	3.36	3	5.5
Overlap	0.58	0	60.2
Common-sample decomposition: 809 issuers			
Connected-only	8.36	9	0.0
Named-only	4.25	3	0.0
Overlap	1.48	1	0.0

Panel B: Standardized signals on the decomposition sample

Portfolio	Return SD (%)	Effect (pp)	SE (pp)
Connected-only	11.75	2.296***	0.373
Named-only	12.21	2.587***	0.348
Overlap	17.55	2.529***	0.265
Issuers / issuer-months	809 / 26,020		

Notes: Panel A counts fixed portfolio members once per issuer; structural zeros are included in the broad sample. The broad sample is Table 15, column (2); the decomposition sample is Table C.2. Panel B divides each disjoint group return by its pooled sample standard deviation across the same 26,020 issuer-months ($n-1$ denominator). The outcome and SIC2 return are unchanged. Effects and standard errors are displayed in issuer-return percentage points per one-standard-deviation increase in the group return. All three signals enter jointly. Scaling leaves fitted values, R^2, and p-values unchanged. Standard errors are two-way clustered by issuer and calendar month. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

C.3 Pairwise Business Similarity and Common Exposures

Table C.4 estimates the following specification separately for each outcome and horizon:

$$\rho_{ij,w} = \beta z(s_{ij}) + \alpha_i + \delta_{j,w} + \Gamma' X_{ij} + \varepsilon_{ij,w}, \tag{14}$$

Here s_{ij} is pre-IPO cosine similarity, w identifies identical calendar start and end dates, and X_{ij} contains the pair controls listed in the table. Dyadic standard errors allow dependence

among pairs sharing either firm, including firms in opposite roles (Aronow et al., 2015). Sample-specific standardization and coverage limit comparisons of coefficient magnitudes across pools. The quarterly operating correlations provide supporting evidence given the short accounting histories.

Table C.4: Business Similarity and Common Economic Exposures

	Expanded pool		Top 30		Same SIC4	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Daily return correlations						
Post-IPO months	2–12	13–36	2–12	13–36	2–12	13–36
Raw returns	0.0094***	0.0122***	0.0087***	0.0120***	0.0033**	0.0082***
	(0.0008)	(0.0010)	(0.0014)	(0.0016)	(0.0016)	(0.0020)
FF5 + momentum	0.0069***	0.0091***	0.0089***	0.0125***	0.0036**	0.0097***
	(0.0007)	(0.0007)	(0.0015)	(0.0017)	(0.0018)	(0.0022)
FF5 + momentum + industry	0.0060***	0.0078***	0.0077***	0.0112***	0.0039**	0.0092***
	(0.0006)	(0.0007)	(0.0015)	(0.0016)	(0.0019)	(0.0022)
Pairs: raw / FF5 + momentum	339,570	259,966	93,747	71,823	43,478	33,911
IPOs: raw / FF5 + momentum	3,409	3,032	3,409	3,032	2,794	2,471
Pairs: plus industry	339,024	259,515	93,207	71,378	43,478	33,910
IPOs: plus industry	3,390	3,014	3,390	3,014	2,794	2,470
Panel B: Quarterly operating correlations						
	Unadjusted	Industry adjusted	Unadjusted	Industry adjusted	Unadjusted	Industry adjusted
Sales growth	0.0228***	0.0249***	0.0214**	0.0169*	0.0218*	0.0250**
	(0.0041)	(0.0045)	(0.0087)	(0.0102)	(0.0114)	(0.0124)
Pairs	165,196	164,554	44,862	44,244	18,963	18,955
IPOs	2,555	2,507	2,540	2,493	1,890	1,882
Change in EBITDA/assets	-0.0005	0.0010	0.0336***	0.0264**	-0.0304***	-0.0287**
	(0.0036)	(0.0039)	(0.0104)	(0.0118)	(0.0110)	(0.0113)
Pairs	159,675	159,126	42,429	41,898	23,168	23,162
IPOs	2,530	2,481	2,516	2,468	1,959	1,953

Notes: Entries are changes in unstandardized pairwise correlations per one-standard-deviation increase in cosine similarity, standardized separately in each regression sample before absorbing fixed effects. The expanded pool contains the locked top 30 matches and other eligible same-SIC2 firms at the initial-range cutoff. Columns restrict this pool as indicated. Every model includes issuer and public-firm-by-aligned-calendar-window fixed effects, same-SIC2/SIC4 and named-competitor indicators, and absolute gaps in pre-IPO log assets, profitability, and tangibility. Panel A omits the listing calendar month (month 1) and requires 120/240 common daily observations in months 2–12/13–36. Factor residuals use an intercept, FF5, and momentum on common dates; the additional industry regressor is the equal-weighted return of the issuer's frozen eligible SIC2 pool excluding both firms. Panel B uses the first 12 full calendar quarters after the listing quarter, requiring eight observations aligned by period-end month. Sales growth and the change in quarterly EBITDA/assets are year-over-year measures, winsorized at pooled unique-firm-month 1st/99th percentiles before pairing. EBITDA is operating income before depreciation. Sales and assets require positive lagged bases; profitability also requires positive current assets. Annual lags require unchanged currency and matching fiscal quarters in consecutive fiscal years. Industry adjustment removes the corresponding leave-pair-out SIC2 fundamental series. Samples require observed outcomes and controls, without imputation; counts differ across specifications. Parentheses report dyadic standard errors allowing shared firms in either role. Two-sided normal tests are unadjusted for multiple comparisons and are outside the principal economic-validation family in Table F.1. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

D Risk

D.1 Issuer Idiosyncratic Volatility

The main risk tests in Section 7 use issuer total volatility as the outcome. Table D.1 reports the complementary specification using issuer IVOL, estimated from the Fama–French five-factor model augmented with momentum separately within each post-listing window. Connected and industry IVOL enter as standardized logs, and the controls, fixed effects, and CR2/Satterthwaite inference follow the main pre-offer specification.

Table D.1: Connected Stock IVOL and Issuer Post-IPO IVOL

	Days 2–20	Days 21–63	Days 64–126	Days 127–252
Connected stock IVOL	0.103***	0.082***	0.076***	0.059***
	(0.018)	(0.013)	(0.011)	(0.013)
SIC2 industry benchmark	0.001	-0.053**	-0.030	-0.002
	(0.035)	(0.022)	(0.022)	(0.016)
IPO controls	Yes	Yes	Yes	Yes
SIC2, offer year, exchange FE	Yes	Yes	Yes	Yes
N	3,954	3,951	3,938	3,903
Adjusted R^2	0.467	0.549	0.576	0.590

Notes: IPO-level OLS regressions of log issuer IVOL over the indicated post-listing trading days. IVOL is annualized FF5 plus momentum residual volatility, with factor loadings estimated within each outcome window. Days 2–20 exclude day 1 and require at least 15 valid daily returns; the remaining windows require at least 30. Days are indexed using valid issuer returns on or after the reconciled first-trading date. Connected stock IVOL is the equal-weighted average for the ten closest connected firms over six pre-offer months; it and the SIC2 industry IVOL benchmark enter as standardized logs. The dependent log IVOL and log IVOL regressors are winsorized at the 1st and 99th percentiles. IPO controls are log gross proceeds, log(1 + age), VC or growth-capital backing, an Internet indicator, and offer price. All specifications include SIC2, offer year, and exchange fixed effects. CR2 standard errors in parentheses are two-way clustered by issuer SIC2 and offer month. Stars use CR2/Satterthwaite p-values: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Issuers whose connected firms have higher pre-offer IVOL exhibit higher idiosyncratic volatility after listing. In Table D.1, a one-standard-deviation increase in log connected IVOL predicts approximately 10.9% higher issuer IVOL over trading days 2–20. The corresponding proportional differences in issuer IVOL are 8.6% over days 21–63, 7.9% over days 64–126, and 6.1% over days 127–252. The connected-IVOL coefficient is positive and significant at the 1% level in every window, conditional on industry IVOL, the included issuer and offering characteristics, and the fixed effects.

Table D.2 assesses whether these issuer-IVOL results over days 21–63, 64–126, and 127–252

depend on the connected set size or pre-offer volatility window. Issuers with more volatile connected firms exhibit higher post-listing volatility across all reported choices of connected set size and pre-offer window. The connected-volatility coefficient remains significant at 5% throughout.

In the issuer volatility regressions, log issuer IVOL and the log IVOL regressors are winsorized at the 1st and 99th percentiles within each estimation sample before the regressors are standardized. Each offering enters once. Volatility estimates use one daily return per security and trading date and are matched to offerings using security identifiers and first trading dates. Manual identifier checks correct one conflicting CRSP match.

Table D.2: Connected Stock IVOL and Issuer Post-IPO IVOL: Robustness

	Days 21–63	Days 64–126	Days 127–252
Panel A: connected set size K (six-month window)			
$K = 5$	0.064***	0.056***	0.035***
	(0.012)	(0.008)	(0.010)
$K = 10$	0.082***	0.076***	0.059***
	(0.013)	(0.011)	(0.013)
$K = 15$	0.091***	0.085***	0.066***
	(0.011)	(0.011)	(0.014)
$K = 20$	0.099***	0.090***	0.077***
	(0.013)	(0.012)	(0.014)
$K = 30$	0.112***	0.099***	0.080***
	(0.015)	(0.012)	(0.014)
N, $K = 5$	3,951	3,938	3,903
N, $K \geq 10$	3,951	3,938	3,903
Panel B: alternative pre-offer windows L ($K = 10$)			
$L = 3$ months	0.068***	0.062***	0.051***
	(0.013)	(0.010)	(0.011)
$L = 12$ months	0.093***	0.093***	0.075***
	(0.011)	(0.009)	(0.014)
N, both windows	3,951	3,938	3,903

Notes: Each cell reports the coefficient on standardized log connected stock IVOL in an IPO-level OLS regression of log issuer IVOL over the indicated post-listing trading days. IVOL is annualized FF5 plus momentum residual volatility. Connected stock IVOL is the equal-weighted average for the top K connected firms over L pre-offer months. All regressions retain the six-month SIC2 industry benchmark, IPO controls, and SIC2, offer year, and exchange fixed effects of Table D.1. CR2 standard errors in parentheses are two-way clustered by issuer SIC2 and offer month. Stars use CR2/Satterthwaite p-values: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

D.2 Incremental Information about Signed Returns

Table D.3 relates first-day and 12-, 24-, and 36-month market-adjusted returns to connected IVOL and returns using the controls listed in its notes. The connected-volatility coefficient is not significant at 5% in any of these specifications. Connected returns remain positive and significant at 5% in the first-day specification.

Continuous regression variables in the signed return tests are winsorized at the 1st and 99th percentiles within each estimation sample.

Table D.3: Connected Stock IVOL and IPO Returns

	(1)	(2)	(3)	(4)
	First-day	Market-adjusted buy-and-hold return		
	return	12 months	24 months	36 months
Connected stock IVOL (z)	0.011	0.035	0.019	-0.017
	(0.014)	(0.041)	(0.026)	(0.032)
Connected return	0.236**	-0.075	-0.110	-0.064
	(0.081)	(0.042)	(0.068)	(0.043)
Industry and market controls	Yes	Yes	Yes	Yes
IPO controls	Yes	Yes	Yes	Yes
SIC2, offer year, exchange FE	Yes	Yes	Yes	Yes
N	3,954	3,913	3,861	3,821
Adjusted R^2	0.337	0.069	0.071	0.065

Notes: IPO-level OLS regressions. First-day return is the first close divided by the offer price minus one; market-adjusted buy-and-hold returns begin at that close (Appendix E). Connected stock IVOL and connected return are equal-weighted measures for the ten closest connected firms. IVOL uses six pre-offer months. Directional returns start six calendar months before the offer date and end at the market close immediately before first public trading. IVOL is annualized FF5 plus momentum residual volatility and enters as a standardized log; connected return is a simple return in levels. Controls include SIC2 industry IVOL and return, market volatility and return, and the IPO controls in Table D.1. All columns include SIC2, offer year, and exchange fixed effects. CR2 standard errors in parentheses are two-way clustered by issuer SIC2 and offer month. Stars use CR2/Satterthwaite p-values: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

E Data Construction

Issuer and public firm disclosures. The information cutoff is the EDGAR filing date of the earliest valid initial filing range in the issuer's pre-offer registration statements or, when necessary, preliminary or final 424B prospectuses. Eligible issuer filings are S-1, F-1, SB-1, SB-2, and S-11 registrations and amendments filed within 730 days before the offer, strictly before the offer, and no later than the cutoff. I use the earliest eligible filing containing a substantive Business section. Issuers without a dated EDGAR price range enter only

through an original registration filed before the offer. Only dated EDGAR ranges identify the pre-range and bookbuilding windows; ranges from other sources can enter pricing-level comparisons.

The disclosure candidate pool contains Russell 3000-linked 10-K, 10-KSB, 10-K405, and 10-KT filings and amendments. Within firm and fiscal year, original filings receive priority, followed by the latest filing date. Russell membership within one year of the fiscal year defines the disclosure corpus; trading status is assessed separately at the issuer's information cutoff. Eligible filings have a usable Item 1 embedding, an EDGAR filing year immediately preceding the IPO offer year, and a filing date no later than the issuer's cutoff. When the range date is unavailable, the issuer's original registration filing date supplies the cutoff.[9]

Connected-firm selection. Let s_i denote IPO i's offer date, $y(i)$ its offer year, c_i its information cutoff, and d_i^- the last market session strictly before c_i. Let $\mathcal{D}_i$ contain public firms with usable prior-year filings in the disclosure corpus that were on file by c_i. The eligible comparable set additionally requires a verified historical security identity, active listing status at c_i, and an observed trade at d_i^-:

$$\mathcal{F}_i \equiv \{\, e^K_{f,\,y(i)-1} : f \in \mathcal{D}_i,\ \text{identity verified},\ f \text{ is actively listed at } c_i \text{ and traded at } d_i^- \,\}.$$

Before ranking, I verify historical security identities using dated CUSIP and CRSP security-history records. A comparable must have active listing status at the cutoff and an observed transaction price and positive volume in the strictly preceding market session. I retain the highest-similarity filing within CIK and historical PERMCO and hold the ranking fixed across tests. Each measure uses available observations among the selected firms; unavailable firms are neither replaced nor assigned new ranks. Candidates with unresolved security links or trading status are excluded from ranking but remain in the construction records, as do IPOs without a connected set. Later delisting leaves the original membership unchanged. The transaction requirement uses CRSP's transaction-price flag and positive observed price and volume.

Table E.1 reports coverage by offer year. The point-in-time network covers 4,207 of the 4,619 IPOs in the initial sample (91.1%). Coverage is lowest in 1996, when EDGAR's electronic filing mandate took effect: 380 of the 672 IPOs have a connected set (56.5%).

[9]The embedding model is applied retrospectively; masking identifying information does not eliminate potential look-ahead bias from pretraining.

The financing, pricing, and issuer-IVOL estimates remain similar when I exclude 1996. The connected book and market leverage coefficients are 0.129 and 0.172; the six-month connected return coefficient in the first-day return regression is 0.280; the bookbuilding return coefficient in the offer price revision regression is 0.138; and the connected IVOL coefficient for issuer IVOL over trading days 21–63 is 0.090. The last comparison concerns the residual-volatility outcome in Appendix D. Four coefficients remain statistically significant at the 5% level; book leverage is significant only at 10%.

Table E.1: Point-in-Time Network Coverage by Offer Year

Offer years	IPO sample	IPOs with connected sets	Coverage (%)
1996	672	380	56.5
1997	438	412	94.1
1998–2000	1,042	992	95.2
2001–2007	760	746	98.2
2008–2015	710	700	98.6
2016–2025	997	977	98.0
All	4,619	4,207	91.1

Notes: The IPO sample is defined in Section 2. Coverage is the percentage of issuers in each offer-year group with connected firms identified under the historical filing cutoffs. Appendix E explains the lower coverage during the transition to mandatory electronic filing in 1996.

Illustrative comparisons. Table E.2 presents eight illustrative IPOs from different periods and industries. For each issuer, the table reports the five public firms with the highest textual similarity and summarizes the business activities underlying the matches. The examples show the selected firms before applying measure-specific accounting or return requirements. These examples illustrate how business descriptions identify economically related firms, including those operating across conventional industry boundaries.

Self-matches. I exclude the issuer's own stock from the canonical ranking using CIK, PERMNO, and PERMCO. Accounting aggregation additionally checks issuer PERMNO and GVKEY. These restrictions remove direct identifier-based self-matches from the connected firm benchmarks. They do not systematically exclude parent–subsidiary relationships or other related entities with distinct identifiers.

Table E.2: Complete Set of Illustrative Comparisons

IPO firm	Year	Business	Closest connected firms	Business activities underlying the match
Medline	2025	Medical supplies and distribution	(1) Omnicell; (2) Insulet; (3) Pulmonx; (4) AngioDynamics; (5) Henry Schein	(1) Medical supplies and medication-management equipment; (2) Insulin-delivery devices; (3) Devices for emphysema treatment; (4) Vascular and cancer-treatment devices; (5) Medical and dental supplies distribution
Kenvue	2023	Consumer health and personal care	(1) The Honest Company; (2) Prestige Consumer Healthcare; (3) Thorne HealthTech; (4) Pacira BioSciences; (5) Patterson Companies	(1) Personal-care and household products; (2) Consumer health and personal-care products; (3) Consumer health products and nutritional supplements; (4) Non-opioid pain treatments; (5) Dental and animal-health supplies distribution
Synchrony Financial	2014	Consumer credit and retail financing	(1) Mastercard; (2) Discover Financial Services; (3) Heartland Payment Systems; (4) American Express; (5) Higher One Holdings	(1) Electronic payment networks; (2, 4) Consumer credit cards; (3) Retail credit and merchant payment processing; (5) Student refund disbursement and payment services
Facebook	2012	Social networking and online advertising	(1) Marchex; (2) Expedia; (3) Online Resources; (4) Constant Contact; (5) LogMeIn	(1, 4) Online advertising and digital marketing; (2) Online travel marketplace; (3) Online banking software (broad textual match); (5) Remote connectivity and collaboration software
General Motors	2010	Automobiles and automotive parts	(1) Tenneco; (2) Navistar International; (3) Commercial Vehicle Group; (4) Dealertrack Holdings; (5) Hughes Communications	(1) Vehicle production and component supply; (2) Vehicle and engine manufacturing; (3) Commercial-vehicle components; (4) Automotive retail software and financing services; (5) Satellite communications and automotive telematics

Text extraction and embeddings. I extract issuer Business sections from the primary EDGAR documents, excluding SEC headers, HTML wrappers, table of contents fragments, Risk Factors text, and financial statement indexes. Standalone section headings define the extraction boundaries. Public firm Item 1 text ends at the next recognized heading, including Item 1A, Item 1B, or Item 2. Combined Business and Properties sections retain the Properties text; Item 1 text incorporated by reference is obtained from Exhibit 13 when available.

Before embedding, I replace company names and aliases, tickers, brands, proprietary

Table E.2: Complete Set of Illustrative Comparisons (continued)

IPO firm	Year	Business	Closest connected firms	Business activities underlying the match
Visa	2008	Electronic payments network	(1) MasterCard; (2) American Express; (3) Online Resources; (4) Net 1 UEPS Technologies; (5) Western Union	(1) Payment networks across SIC2 codes; (2) Card-payment networks; (3) Electronic payments and bill payment; (4) Smart-card payment processing; (5) Money transfer and consumer payment services
United Parcel Service	1999	Package delivery and logistics	(1) FDX; (2) AirNet Systems; (3) Expeditors International of Washington; (4) Ingram Micro; (5) Black Box	(1) Parcel delivery and logistics; (2) Time-critical air and ground delivery; (3) Freight transportation and logistics; (4) IT distribution and order fulfillment; (5) Network equipment distribution and technical services
Conoco	1998	Oil and gas production and refining	(1) Chevron; (2) Occidental Petroleum; (3) Unocal; (4) Snyder Oil; (5) Mobil	(1) Oil and gas production, refining, and marketing; (2, 3, 4) Oil and gas exploration and production; (5) Oil and gas production, refining, and chemicals

Notes: The table lists the five closest public firms for eight illustrative IPOs from different periods and industries. Firms are ordered by cosine similarity of business descriptions; numbers link each firm to its activities. Table 2 presents four of these examples. Appendix E describes selection.

product names, person names, and web addresses with category placeholders. Generic products, business activities, and text outside the identified spans remain unchanged. The eligible universe excludes fifteen public filings for which the masking procedure produces no usable result. The corpus contains 4,227 issuer descriptions and 82,348 public descriptions. I split long masked documents into nonoverlapping 7,500-token chunks, embed each chunk with text-embedding-3-small, and average the chunk vectors using token count weights. I normalize each document vector to unit Euclidean norm. The resulting vectors have 1,536 dimensions, and the same masked ranking supplies all connected-set measures. Missing descriptions do not enter the similarity match.

Alternative comparable sets. The industry benchmarks contain eligible public firms in the issuer's two-digit or four-digit SIC industry. The explanatory-power comparison in Table 3 selects ten characteristic-matched firms across industries using standardized log assets, profitability, and tangibility. Its portfolios equally weight available constituents; sets can overlap, and each pairwise comparison uses its own common sample. The embedding set contains the ten closest firms selected in Section 3. The monthly comparison retains partial

return histories and observed constituent returns without market reinvestment.

TF-IDF uses lowercase word unigrams, English stopword removal, and raw term frequencies. At each IPO cutoff, terms must occur in at least ten eligible public filings. I then select up to 1,000 terms by aggregate raw token count, resolving ties alphabetically. Inverse document frequency is $1 + \log((1 + N)/(1 + df))$, where N is the number of eligible public filings and df is document frequency. L2 normalization and cosine similarity determine the ten closest historical companies. Vocabulary and inverse document frequencies use only eligible public filings available by the cutoff.

In Table 8, the size-matched set contains the ten public firms in the issuer's two-digit SIC industry closest in log market capitalization. Public-firm capitalization is measured on the last trading day before the range filing; expected issuer capitalization is the initial-range midpoint times shares outstanding at listing. The characteristic-matched set contains ten firms selected across industries by pre-IPO log assets, profitability, and tangibility from the market-leverage panel. The SIC4 portfolio requires at least three firms with returns; the size- and characteristic-matched portfolios require at least five. Each set may contain connected firms. All portfolios share the pre-range and bookbuilding windows, while the duration control measures the full range-to-trading period. Each regression uses observations with all required variables. The financing comparison in Table 11 instead uses five characteristic-matched firms.

Accounting dates and missing debt. Issuer leverage uses the first Compustat quarter-end on or after the offer, within 270 days, without an annual fallback. Pre-IPO leverage uses the last quarter-end strictly before the offer, and leverage one and two years after listing uses the quarter-end closest to the listing statement date plus 365 and 730 days, within 60 days. Issuer characteristics use the latest annual fiscal period end strictly before the offer. Connected firm accounting variables use the latest annual fiscal period end no later than the matched public filing date and the offer date; where filing dates are absent, the accounting fiscal year must match the linked filing's fiscal year. Market equity uses the closest CRSP month-end within 45 days of the accounting date.

Total debt is current plus long-term debt. The baseline construction treats a single missing component as zero but leaves total debt missing when both components are absent; the complete component sensitivity check removes reliance on that convention. Ratios require positive denominators. Debt change is the difference in balance sheet debt between the last

quarter-end before the offer and the first quarter-end on or after it, divided by pre-offer assets. It is not a financing cash flow.

The leverage regressions winsorize continuous regression variables at the 1st and 99th percentiles of the merged IPO sample before selecting the complete-case estimation sample. Connected firm and SIC2 industry leverage are constructed using the fiscal-year winsorization described in Section 5.

In Table 13, the debt-change ratio and pre-IPO leverage use 1st and 99th percentile cutoffs computed across the matched IPO sample before complete-case restrictions. Leverage changes subtract the separately winsorized leverage levels, without further winsorization of the difference. The distances above and below each benchmark are not winsorized again. Table 14 instead winsorizes each debt-change outcome separately at the 1st and 99th percentiles of its common accounting-history cohort; consequently, its coefficients need not add across periods.

Debt-change timing. The common sample in Table 14 requires observed debt components and currencies matching the pre-IPO quarter. The initial-leverage spline uses cohort-specific knots held constant across outcomes, with additional fixed effects for five-percentage-point intervals of starting leverage. Quarter -1 is the last pre-offer fiscal quarter-end, quarter 0 is the next scheduled quarter-end on or after the offer, and quarter $+4$ is four fiscal quarters later. With D_q denoting debt in quarter q and A_{-1} pre-IPO assets, the preceding, IPO, and post-listing changes are $(D_{-1} - D_{-2})/A_{-1}$, $(D_0 - D_{-1})/A_{-1}$, and $(D_4 - D_0)/A_{-1}$. The IPO-minus-preceding-quarter contrast is estimated directly from $(D_0 - 2D_{-1} + D_{-2})/A_{-1}$. Each outcome is winsorized separately, so estimated coefficients need not add across periods. Table A.10 also requires the earlier debt observation and reports $(D_{-2} - D_{-3})/A_{-1}$ and $(D_0 - D_{-3})/A_{-1}$.

Post-IPO buy-and-hold returns. Issuer performance in the signed-return tests of Appendix D is measured from the first-day close t_0 to the same calendar day H months later. The raw buy-and-hold return compounds the issuer's daily CRSP total returns over that window,

$$\mathrm{BHR}_{i,H} = \prod_{t \in (t_0,\, t_0+H]} (1 + R_{i,t}) - 1, \tag{15}$$

where $R_{i,t}$ is the issuer's daily total return, with delisting treatment described below. The market-adjusted return is

$$\mathrm{BHAR}_{i,H}^{IPO} = \prod_{t}(1 + R_{i,t}) - \prod_{t}(1 + R_{m,t}), \tag{16}$$

where $R_{m,t}$ is the Fama–French daily market return, defined as the market excess return plus the risk-free rate, compounded through the nominal horizon.

Monthly returns in the comovement regressions of Section 6 compound the same daily total returns within each calendar month. The issuer's delisting month includes its delisting return; later months are excluded. A delisted connected firm earns the market return from its delisting day, and the portfolio averages over firms with an observed return in the month.

The connected portfolios in Tables 15 and 16 contain firms meeting the six-month pre-offer return requirement of at least 60% of expected trading days. Similarity ranks are assigned before this screen and remain unchanged. Each rank-band regression requires all three portfolio returns; column (1) of Table 16 also uses the complete-case sample of the finer four-band partition. The matched weighting comparison requires positive capitalization in the preceding calendar month and omits months after constituent delisting. Value weights normalize lagged capitalization within each band and month. The monthly SIC2 benchmark excludes the issuer and requires at least five firms.

Return endpoints and delisting. Daily and monthly returns are obtained from CRSP for U.S.-incorporated common equities. I incorporate the latest delisting event for each PERMNO when it falls within ten calendar days of the last ordinary daily return, using the CRSP 2.0 event return and the legacy daily delisting return as a fallback. I apply the delisting return once, on the day after the later of the delisting date and the last ordinary trading date. For performance-related delistings (codes 500 and 520–584) with missing delisting returns, I assign -30% on the NYSE and AMEX and -55% on Nasdaq, following Shumway (1997) and Shumway and Warther (1999). Other delistings without a recorded return are measured through the last ordinary trade.

For long-horizon return calculations, proceeds after delisting are reinvested in the market through the specified horizon, without a minimum survival requirement. In the monthly-return robustness analysis, I replace the supplied delisting-month return by compounding ordinary daily returns through delisting and applying the delisting return once. Industry benchmarks use CRSP monthly returns without an additional delisting adjustment.

Named competitors. I identify named competitors from the selected pre-IPO registration statement, including the Business and Risk Factors sections. Disclosed names are linked to historical public-firm identities using security identifiers and pre-IPO trading records. The set contains 9,826 issuer–company links across 2,508 IPO records before return-availability requirements. I compare these named competitors with the text-based connected firms to examine whether the two approaches identify overlapping or complementary business relationships. Competitor coverage is incomplete, and eligibility differs from that of the text-based set. The underlying source records preserve unresolved classifications.

Winsorization and bootstrap inference. The first-day and pricing regressions use 1st and 99th percentile cutoffs computed on the full IPO panel for continuous variables. Appendix F describes coefficient and contrast inference and the adjustment-share bootstrap.

Sample selection. The initial sample contains 4,619 IPOs. Substantive business descriptions are available for 4,227 issuers, and 4,207 have connected sets, six-month connected returns, and baseline connected IVOL. Initial filing ranges are observed for 4,331 IPOs and range filing dates for 3,893; connected bookbuilding returns are available for 3,850. Accounting data identify 4,237 issuers, of whom 4,210 have observed book leverage. These counts precede regression controls and outcome requirements.

Each analysis uses a separate fixed sample. The first-day pricing sample contains 3,954 IPOs with the baseline six-month signal, initial filing range, and complete controls. Monthly comovement and baseline pre-offer volatility tests use this sample, subject to their additional outcome requirements. The dated bookbuilding sample contains 3,830 complete cases at the $T-2$ cutoff. The book and market leverage samples contain 3,728 and 3,726 IPOs, respectively, under the preferred listing specifications and sector restrictions. Table 1 reports available observations within these samples, as specified in its note; continuous variables are winsorized at the 1st and 99th percentiles. I do not restrict the analyses to their common intersection.

The alternative leverage comparisons retain all 3,728 book-leverage and 3,726 market-leverage issuers for both the SIC2 industry and characteristic-matched benchmarks (Table 11). Additional data requirements reduce particular estimation samples: requiring both debt endpoints and issuer characteristics yields 3,699 (Table A.3). The five-firm bookbuilding specification uses 3,830 IPOs (Table B.3). The post-listing comovement panel contains

127,963 issuer-months for 3,942 issuers (Table 15).

Within the comovement sample, complete cases for the connected-only and all named-firm comparison yield 2,358 issuers and 76,145 issuer-months. The symmetric comparison contains 2,228 issuers and 71,840 issuer-months, and the three-group model contains 809 issuers and 26,020 issuer-months. All follow the return construction in Section 6 without replacing unavailable group returns.

Volatility specifications additionally require estimable factor residuals and the applicable return or pricing outcome; each table reports its resulting sample. The issuer volatility panel contains one observation per offering.

E.1 Aggregation Weights

I compare equal weights with weights proportional to the original pre-IPO cosine similarity scores. Let $\mathcal{C}_i = \mathcal{C}_i(10)$ denote issuer i's original connected set, s_{ij} its similarity to firm j, and $\mathcal{A}_i^X \subseteq \mathcal{C}_i$ the firms satisfying the existing eligibility requirements for characteristic X. The full-set base weights and ordinary connected averages are

$$w_{ij}^0 = \frac{s_{ij}}{\sum_{k\in\mathcal{C}_i} s_{ik}}, \qquad X_i^{CW} = \frac{\sum_{j\in\mathcal{A}_i^X} s_{ij} X_j}{\sum_{j\in\mathcal{A}_i^X} s_{ij}}.$$

Each comparison retains the same constituent observations and issuer sample as its equal-weighted counterpart. All original selected scores are positive and finite. Unavailable firms are not replaced by lower-ranked firms.

Pre-IPO returns average the existing individual-firm cumulative returns, retaining the original endpoints, including $T-2$ for bookbuilding. Monthly returns use the original scores and the baseline availability and delisting conventions. Leverage averages firm-level ratios; connected accounting controls use the same weighting rule, and leverage gaps are recalculated after the existing transformations. Post-listing leverage benchmarks remain fixed at their original pre-IPO dates. Total volatility and IVOL average eligible individual-firm estimates before taking logs and applying the existing winsorization and standardization. Announcement news instead uses fixed exposure weights $m_i w_{ij}^0$, where m_i is the original set's total assigned exposure. These weights preserve that mass and the existing distinction between zero exposure and unavailable contributions, without renormalizing to announcing firms or measured contributions.

Table E.3 compares the nine principal tests defined in Appendix F. Each coefficient or contrast changes by less than 1% in magnitude, and conclusions under the existing within-family Holm adjustments are unchanged. An 11.8-percentage-point increase in connected leverage is associated with year +1 issuer leverage higher by 2.04 percentage points under equal weights and 2.05 percentage points under cosine weights. Conditional listing leverage, below-gap debt adjustment, and the announcement-news main effect remain imprecise; the connected-IVOL specifications and other volatility horizons also remain similar. Changes in cross-sectional adjusted R^2 are below 0.0003.

The weights are close to equal weights. Among the 4,205 complete original ten-firm sets, the median largest weight is 10.29%, and its 95th percentile is 10.84%. The median effective number of connected firms, $1/\sum_j (w_{ij}^0)^2$, is 9.9979, with a minimum of 9.9272. These statistics describe complete matching sets; measure-specific eligible sets can be smaller, occasionally containing one firm. Across the matched analysis samples, correlations between the corresponding signals before regression preprocessing exceed 0.9998. Eight of the 3,697 issuers in the debt-adjustment sample change sides of the leverage benchmark (0.216%). The comparison documents insensitivity to raw-cosine weighting within the selected sets; the small signal differences limit what it establishes about incremental information in similarity intensity.

The comparison uses the principal bootstrap margins. Appendix F's alternative-margin qualifications remain relevant: the equal-weighted other-margin p-values are 0.0505 for the financing common-slope contrast and 0.0604 for year +1 leverage. Cosine-weighted tests under those alternative margins are not reported.

Table E.3: Principal Results under Equal and Cosine-Similarity Weights

Principal coefficient or contrast	Equal weights	Cosine weights	Holm p-value EW / CW	Common N
Pricing				
Bookbuilding return: offer-price revision	0.1332	0.1338	0.0120 / 0.0117	3,830
Bookbuilding return: first-day return	0.3675	0.3691	0.0024 / 0.0024	3,830
Bookbuilding return: first-day return, conditional on revision	0.2990	0.3002	0.0120 / 0.0117	3,830
Announcement news × uncertainty: total repricing	1.1649	1.1722	0.0120 / 0.0117	3,632
Financing				
Above-gap debt adjustment	−0.2523	−0.2508	0.0306 / 0.0345	3,697
Common-slope contrast, $\beta_A + \beta_B$	−0.2823	−0.2809	0.0306 / 0.0345	3,697
Year +1 leverage, conditional on pre-IPO leverage	0.1728	0.1740	0.0359 / 0.0360	3,514
Economic validation				
Monthly return comovement	0.3303	0.3311	< 0.0002 / < 0.0002	127,963
Connected total volatility: issuer total volatility, days 21–63	0.0878	0.0881	0.0002 / 0.0002	3,951

Notes: Each row compares the indicated coefficient or contrast under equal (EW) and raw-cosine-proportional (CW) connected-firm weights. Constituent observations, issuer samples, information dates, eligibility requirements, specifications, fixed effects, conventional benchmarks, and preprocessing policies are identical within each comparison. Connected accounting controls and leverage gaps are reconstructed consistently. Reported p-values are Holm-adjusted within the existing four-test pricing, three-test financing, and two-test economic-validation families in Appendix F. The underlying two-sided tests use the null-imposed wild cluster bootstrap with two-way CR1 studentization, 9,999 requested draws, seed 20261114, and the principal bootstrap margins specified there. Comovement has zero bootstrap exceedances; its adjusted p-values are reported below twice the simulation resolution. The volatility coefficient is per one-standard-deviation increase in the winsorized log connected-volatility signal. Monthly observations are issuer-months for 3,942 issuers. The financing common-slope restriction is $\beta_A + \beta_B = 0$. The comparison preserves the existing alternative-margin qualifications in Table F.1; CW alternative-margin tests are not reported.

F Statistical Inference

F.1 Estimation Samples and Test Families

All inference procedures hold the regression specifications, transformed variables, and estimation samples fixed, so they compare uncertainty around the same OLS estimates. The financing models also retain the leverage spline and its knots.

Table F.1 compares the null-imposed bootstrap tests reported in the main tables with analytic and product-weight bootstrap tests. The pricing family comprises the three bookbuilding-return coefficients in Table 7—offer price revision and first-day return without and with revision conditioning—and the total-price uncertainty interaction in Table 9. The financing family comprises the observed-benchmark listing-quarter above-gap debt coefficient and common-slope contrast in Table 13, Panel A, and the conditional year +1 connected-leverage coefficient in Table 12. The economic-validation family comprises the controlled monthly-comovement coefficient in Table 15 and the days 21–63 connected-total-volatility coefficient in Table 17, Panel A. Raw and Holm-adjusted p-values are reported within these four-, three- and two-test families. The five-test financing sensitivity additionally includes the two predicted-gap checks in Panel B of Table 13.

For these coefficients, and for the listing-quarter above-gap coefficient and slope sum after adding predicted gaps, the main tables report null-imposed bootstrap p-values; any stars and confidence intervals for the same tests use that procedure. Stars use raw, two-sided p-values. Standard errors in parentheses remain CR2 estimates for the cross-sectional regressions and two-way clustered panel estimates for monthly comovement. Other specifications use the analytic tests and, where reported, product-weight bootstrap described below; they do not enter the three Holm families. Each table identifies the method used.

F.2 Multiway Clustering and Wild Bootstrap

Errors are clustered by SIC2 industry and offer month for pricing and volatility, SIC2 industry and offer year for financing, and issuer and calendar month for monthly comovement. The two-way covariance estimator combines the two clustering dimensions and subtracts their intersections:

$$\widehat{V} = \widehat{V}_1 + \widehat{V}_2 - \widehat{V}_{12}.$$

Each component includes finite-sample corrections that account for the number of clusters and estimated parameters, including absorbed fixed effects (Cameron et al., 2011).

The null-imposed wild cluster bootstrap applies Rademacher weights to restricted residuals along one clustering dimension and retains both dimensions in CR1 studentization (Roodman et al., 2019). Weights are assigned by SIC2 for pricing and volatility, offer year for financing, and calendar month for comovement. Each choice uses the dimension with fewer clusters; Table F.1 reports the cluster counts. Tests using the other weighting dimension assess sensitivity to this choice. Issuer effects are reabsorbed within each monthly-comovement replication.

Each two-sided test requests 9,999 replications, and the 95% confidence sets invert the same tests. The p-values use replications with valid studentizing variances; zero exceedances are reported as bounds at the simulation resolution. At the zero null, undefined variances occur in at most 0.130% of replications under the reported weighting dimension (Table F.1).

F.3 Analytic Comparisons and Cluster Sensitivity

The cross-sectional analytic comparisons use a two-way CR2 calculation with coefficient-specific Satterthwaite degrees of freedom. The degrees of freedom match the first two moments of the contrast variance under an iid unit-variance working model, so the resulting tests depend on that approximation. Monthly-comovement comparisons use two-way clustered panel-model inference.

The reference wild bootstrap for the cross-sectional tables multiplies restricted residuals by products of independent industry and time weights, uses two-way CR1 studentization, requests 4,999 draws, and calculates $(1 + \text{exceedances})/(1 + B_u)$ over feasible draws.

For Table B.1, columns (1) and (2), the SIC2-weight bootstrap with two-way CR1 studentization gives $p = 0.0003$ under both equal and cosine-similarity weights, compared with product-weight p-values of 0.050 and 0.049, respectively. These six-month first-day return tests are outside the four-test pricing family.

To assess sensitivity to individual clusters, I re-estimate the revision-conditioned first-day coefficient, total-price uncertainty interaction, and observed-benchmark listing-quarter above-gap debt coefficient after omitting each SIC2 industry and each offer year in turn. Transformed regressors, normalizations, spline knots, and economic-effect scales remain fixed.

Every omission preserves the signs of all three coefficients. Across industry and year

omissions, a one-standard-deviation increase in connected bookbuilding returns is associated with a first-day closing price approximately 1.94% to 3.97% higher relative to the offer price, conditional on the revision. For a one-standard-deviation increase in announcement news, the additional percentage price response associated with one-standard-deviation higher uncertainty ranges from approximately 1.89 to 2.89 percentage points. The above-gap debt effect ranges from -3.49% to -2.16% of pre-IPO assets.

F.4 Sensitivity to Bootstrap Margins and Multiple Testing

The four pricing coefficients remain significant at 5% after the within-family Holm adjustment. The adjusted p-values for the revision, unconditional first-day, conditional first-day, and total-repricing uncertainty coefficients are 0.0120, 0.0024, 0.0120, and 0.0120, respectively. All four also remain significant at 5% when offer-month weights replace SIC2 weights. For monthly comovement, the raw p-value is below 0.0001 under either bootstrap margin and the Holm-adjusted p-value is below 0.0002. For days 21–63 total volatility, the raw and Holm-adjusted p-values are 0.0002; offer-month weights give $p < 0.0001$.

The financing results are more sensitive to these choices. With offer-year weights, the observed-benchmark above-gap debt coefficient and common-slope contrast have raw p-values of 0.0102 and 0.0109 and three-test Holm-adjusted p-values of 0.0306. SIC2 weights give 0.0274 and 0.0505, respectively. Adding the two predicted-gap checks to the Holm family raises both adjusted p-values to 0.0510. The conditional year +1 leverage coefficient has a raw and three-test Holm-adjusted p-value of 0.0359, compared with 0.0604 under SIC2 weights and 0.0693 under the five-test correction. Thus the evidence against a common financing slope and for the one-year leverage relation crosses the 5% threshold under alternative procedures. These changes concern precision; the specifications, point estimates, and economic magnitudes are identical.

The analytic and product-weight bootstrap columns of Table F.1 reconcile inference for these same estimates. For the revision-conditioned first-day coefficient, the product-weight bootstrap gives $p = 0.0428$, compared with 0.0040 under the null-imposed SIC2-weight bootstrap. The total-repricing uncertainty interaction has a CR2/Satterthwaite p-value of 0.0178 and a null-imposed bootstrap p-value of 0.0040. For conditional year +1 leverage, the product-weight bootstrap gives 0.0662, compared with 0.0359 under the null-imposed offer-year bootstrap. The procedures differ in residual weighting, studentization, finite-sample

corrections, and calibration, as specified above. For the bookbuilding coefficients in Panels B and C of Figure 1, the analytic 95% intervals are [0.048, 0.218] and [0.045, 0.553]; the figure caption reports the null-imposed bootstrap intervals.

For the subsequent leverage changes in Table 13, Panel B, the one-year above-gap coefficient is significant at 5% under Satterthwaite tests and at 10% under the product-weight bootstrap; the two-year coefficient is significant at 10% under both. Common-slope tests at both horizons reject at 5% under Satterthwaite tests and at 10% under the product-weight bootstrap. In the timing comparison in Table 14, the IPO-minus-preceding-quarter contrast for above-benchmark issuers gives $p = 0.041$ under Satterthwaite tests and 0.055 under the product-weight bootstrap.

F.5 Financing Benchmarks and Other Procedures

Observed connected and industry leverage in Table 13, Panel A, form the financing baseline. A higher connected benchmark reduces the above-benchmark gap and increases the below-benchmark gap, so the common-slope restriction is $\beta_A + \beta_B = 0$. The tests of the above-gap coefficient and this contrast are reported separately. Panel B adds the two predicted-leverage gaps as a conditional robustness check on the same sample. Its inference treats the fitted benchmark as fixed and excludes first-stage estimation uncertainty.

The explanatory-power comparisons in Table 3 use the following procedures. Panels A–C and E use 2,000 unclustered paired studentized IPO-bootstrap draws, conditional on portfolios and transformations, with continuous variables winsorized at the 1st and 99th percentiles within each pairwise sample. Panel D equally weights calendar-month cross-sectional R^2 over the first 36 complete post-listing months. Its paired tests use Newey–West covariance with Bartlett weights, six calendar-month lags and the $N/(N-1)$ correction. These comparisons are unadjusted for multiple testing and are outside the principal test families.

The adjustment-share interval in Table 5 uses a separate bootstrap. Each of 2,000 draws independently resamples SIC2 industries and offer months, weights IPOs by the product of their cluster multiplicities, and refits the revision and first-day equations with common weights. The basic interval reflects the 2.5th and 97.5th percentiles of the bootstrap revision-share ratios around the point estimate. This resampling interval is distinct from the null-imposed wild-bootstrap confidence sets for the principal coefficients.

Table F.1: Inference and concentration of identifying variation

Panel A: Point estimates and inference by method Coefficient or contrast	Estimate	Effect	Reference p: A/W	Bootstrap p	95% bootstrap set	Holm p	Other margin p
Pricing							
Bookbuilding: revision	0.133	1.53	0.0087 / 0.0106	0.0094	[0.043, 0.229]	0.0120	0.0073
Bookbuilding: first-day	0.367	4.23	0.0248 / 0.0363	0.0006	[0.110, 0.586]	0.0024	0.0038
Bookbuilding: first-day conditional on revision	0.299	3.44	0.0282 / 0.0428	0.0040	[0.077, 0.481]	0.0120	0.0088
Total-price uncertainty interaction	1.165	2.27	0.0178 / —	0.0040	[0.551, 2.279]	0.0120	0.0066
Financing							
Listing debt: above-gap, observed benchmarks	-0.252	-2.98	0.0188 / 0.0286	0.0102	[-0.420, -0.070]	0.0306 [0.0510]	0.0274
Listing debt: common-slope contrast	-0.282	-3.33	0.0265 / 0.0248	0.0109	[-0.502, -0.077]	0.0306 [0.0510]	0.0505
Year +1 connected leverage	0.173	2.04	0.0462 / 0.0662	0.0359	[0.013, 0.315]	0.0359 [0.0693]	0.0604
Listing debt: above-gap, adding predicted gaps	-0.231	-2.73	0.0192 / 0.0304	0.0136	[-0.389, -0.057]	— [0.0510]	0.0285
Listing debt: slope sum, adding predicted gaps	-0.232	-2.73	0.0384 / 0.0355	0.0346	[-0.427, -0.024]	— [0.0693]	0.0422
Economic validation							
Monthly connected-return comovement	0.330	3.30	< 0.0001 / —	< 0.0001	[0.284, 0.376]	< 0.0002	< 0.0001
Issuer total volatility: days 21–63	0.088	9.18	< 0.0001 / —	0.0002	[0.065, 0.112]	0.0002	< 0.0001

Panel B: Clusters and concentration Coefficient or contrast	N	G_1/G_2	ν_{CR2}	df_{CR1}	Undefined var. (%)	Leave SIC2 out	Leave year out
Bookbuilding: revision	3,830	69/337	5.92	68	0.000	—	—
Bookbuilding: first-day	3,830	69/337	5.92	68	0.000	—	—
Bookbuilding: first-day conditional on revision	3,830	69/337	5.92	68	0.000	[0.195, 0.330]	[0.168, 0.345]
Total-price uncertainty interaction	3,632	69/337	6.08	68	0.000	[1.058, 1.482]	[0.968, 1.372]
Listing debt: above-gap, observed benchmarks	3,697	59/30	6.79	29	0.000	[-0.276, -0.187]	[-0.296, -0.183]
Listing debt: common-slope contrast	3,697	59/30	5.05	29	0.000	—	—
Year +1 connected leverage	3,514	59/30	5.91	29	0.000	—	—
Listing debt: above-gap, adding predicted gaps	3,697	59/30	6.77	29	0.000	—	—
Listing debt: slope sum, adding predicted gaps	3,697	59/30	5.13	29	0.130	—	—
Monthly connected-return comovement	127,963	3942/356	—	355	0.000	—	—
Issuer total volatility: days 21–63	3,951	69/338	10.51	68	0.000	—	—

Notes: Estimates, samples, transformations and economic-effect scales are identical across procedures. Reference p: A/W denotes analytic/product-weight wild-bootstrap inference; A uses CR2/Satterthwaite except for comovement's panel-model test. Bootstrap tests use null-imposed Rademacher weights, two-way CR1 studentization, and inverted 95% confidence sets. Error margins are SIC2/offer month for pricing and volatility, SIC2/offer year for financing, and issuer/calendar month for comovement. Bootstrap weights use SIC2, offer year and calendar month, respectively; the other margin provides the sensitivity check. Holm families contain four pricing, three financing and two validation tests; bracketed financing values report the five-test sensitivity including the predicted-gap checks. Appendix F describes the procedures and sensitivities. Pricing effects are 100 times the relevant coefficient and return SD, approximating percentage changes in the price ratios for bookbuilding and the additional percentage-point response for the uncertainty interaction. Other effects are percentages of pre-IPO assets for debt, leverage percentage points for year +1, issuer-return percentage points per ten-point connected return for comovement, and $100[\exp(\beta) - 1]$ for volatility. Financing uses the common 0.1180 leverage SD. ν_{CR2} is the CR2 working-model approximation; $df_{CR1} = \min(G_1, G_2) - 1$ is an analytic reference. Undefined-variance fractions concern the zero null; leave-out ranges cover the three designated influence targets.