What Wins Interest Is Not What Closes the Deal: LLM-Measured Judgment in Early-Stage Finance*

Afroza Alam Müge Süer†

November 13, 2026

Job Market Paper

Abstract

What makes investors interested in an early-stage venture may not be what closes the deal. We build a new dataset of 642 pitches from the U.S. television show *Shark Tank*, following the same ventures from investor interest, through an on-air handshake, to whether that agreement survives due diligence after the show, an outcome no database records. For each pitch, we distinguish realized financial figures (hard information) from three forms of soft information: founders' forward-looking claims, linguistic features captured by conventional textual analysis, and contextual judgment. To measure judgment, we develop an approach based on large language models (LLMs) that reads each pitch in context and forms a judgment of the venture. Judgment is the strongest predictor of competing investor offers, is associated with better terms for founders, and predicts which ventures survive after the show. It remains the strongest predictor at the handshake, although realized and forward-looking financial figures also become predictive at this stage. However, judgment does not predict whether an agreed deal closes. At that stage, founders' forward-looking claims become predictive instead: larger asks, projections and implied valuations are associated with a higher probability that the deal falls through. Different information is predictive at different stages of the same investment process.

JEL codes: C55, G24, L26

Keywords: entrepreneurial finance, early-stage fundraising, large language models, soft information

*We thank Pedro Armada, Karen Bernhardt-Walther, Sena Coskun, Husnu C. Dalgic, Catherine Eckel, Guido Friebel, Jin Li, Javier Miranda, Alexander Reifschneider, as well as the organizers and participants of Experimental Finance Conference 2025, the SEF session at ASSA 2026, OESS 2026, and the AI in Organizations Workshop at Goethe University Frankfurt for their encouragements and helpful comments. We are also very grateful to our research assistant, Maximilian Gottwald, and to our interns, Baris, Lukas, Negar, and Sababa, for their valuable support. Support by the SEF and Chapman University through the Vernon L. Smith Young Talent Award 2024 and the research grant by Behavioral Finance e.V. are gratefully acknowledged.

†Halle Institute for Economic Research (IWH).

1 Introduction

Early-stage financing determines which young firms get the capital to grow (Lerner and Nanda, 2020). Investors fund roughly one of every hundred ventures they consider (Gompers et al., 2020). A large literature studies what makes investors interested in a venture, showing that they rely heavily on soft information about the business and its founders (Kaplan and Strömberg, 2004; Bernstein et al., 2017; Gompers et al., 2020; Hu and Ma, 2025). However, for a founder, getting an investor interested is not the same as being funded. Terms still have to be agreed on, and the agreement still has to survive due diligence, and both of those happen in private. Research usually observes the beginning and the end of this process but little of what connects them. As a result, we know much less about which kinds of information become relevant at each step of the journey from a founder's first exchange with investors to the point where money actually changes hands.

We follow the same ventures through the three decisions of that journey and show that these decisions do not rely on the same information. What wins an investor's interest is not what closes the deal. The strongest predictor of whether a venture draws competing offers is the contextual *judgment* of the business itself, formed by two large language models (LLMs) that read the pitch as a whole. This judgment is also associated with better terms for the founder and predicts which ventures are still active three years later. At the handshake, the agreement reached before due diligence, judgment remains the strongest predictor, but the financial figures also become predictive. Out of sample judgment and the realized figures predict the handshake equally well. Only at completion does judgment lose its predictive power. Instead, the founder's own projections become predictive: larger asks, forecasts, and implied valuations are associated with a lower probability that the deal is completed.

We study this process using a new dataset that we construct from scratch. We transcribe all 642 available pitches from 229 episodes of the U.S. television show *Shark Tank*, check every transcript word by word, and structure each one as a founder–investor dialogue. On air, we observe whether a venture attracts competing offers (*investor interest*) and whether a founder and an investor reach an agreement (*handshake*), as well as the terms of that agreement. What happens after the show is harder to observe. No database records whether a televised handshake was honored or whether a small venture is still operating, so those outcomes survive only in scattered public traces: a later interview, a dead storefront, a state filing. We collect these records as well. Using language-model agents searching the open web twice, independently, we establish for 313 of the agreements whether the deal was completed (*completion*). For the same venture we therefore learn what drew interest, what was agreed, and whether the agreement was completed.

We separate the information in each pitch into four kinds. Realized financial figures describe what the venture has already done (*hard finance*). Forward-looking financial claims, including the ask, describe what the founder says the venture will do (*soft finance*). Conventional textual analysis captures how the founder speaks, through dictionary, sentiment and readability measures (*language*). Finally, two LLMs read the pitch in context and assess features such as product–market

fit, pricing, scalability and execution (*judgment*). Soft finance, language and judgment are all based on soft information, but are transformed into systematic measures in different ways (Liberti and Petersen, 2019). This allows us to compare hard information with different forms of soft information across the same sequence of decisions.

The information associated with investment changes sharply along this journey. First, we show that a standard deviation of judgment raises the probability of drawing competing offers, *investor interest*, by 15.4 percentage points, relative to a base rate of 50.5 percent. The corresponding estimate for hard finance is 4.0 percentage points, whereas the estimates for soft finance and language are close to zero. Judgment does more than attract interest. It is associated with how much ownership founders give up. Most founders receive exactly the money they asked for. But the median founder who offers ten percent of the company leaves having sold twenty, and a standard deviation of judgment is worth 2.7 points of the company, holding the ask, the equity offered and the implied valuation fixed. Judgment also carries information about the venture beyond investor behavior: it raises the probability that a venture is alive three years after airing by 5.1 percentage points, with all four indices and the handshake itself in the regression. This association is at least as strong among ventures that reached no *handshake* as among those that reached one. Linguistic features, by contrast, add little information beyond judgment. On its own, the language index is associated with a 6.2 percentage-point higher probability of interest; controlling for judgment, the estimate falls to 0.5 percentage points and is statistically indistinguishable from zero.

Second, the founder's forward claims, our soft finance index, become increasingly predictive of the *handshake* and of deal *completion*. At the *handshake*, judgment remains the strongest predictor: a standard deviation of judgment is associated with an 8.0 percentage point higher probability of reaching a handshake. Both sets of financial figures also become predictive, with increases of 4.7 percentage points for hard finance and 3.8 percentage points for soft finance. At *completion*, judgment is no longer statistically significant, and only the founder's forward claims remain predictive. A standard deviation of the amount asked is associated with a 10.8 percentage point lower probability that the deal is completed. The corresponding reductions are 12.5 percentage points for projected revenue and 9.1 percentage points for the implied valuation. In other words, the more ambitious the founder's figures about the future, the more often the agreed deal falls through. We do not observe why ambitious claims predict non-completion. Possible explanations include claims that fail verification during due diligence, disagreements over valuation, renegotiation of terms, or founders withdrawing from the deal.

Why does each step reward different information? At *interest*, the question is whether more than one investor wants the venture, and because nothing can be verified on air, what an investor has to go on is an assessment of the business as a whole. A *handshake* requires a founder and an investor to agree on terms (Kaplan and Strömberg, 2001, 2004), and a price is written in the founder's own figures, so those figures necessarily enter. *Completion* requires that the agreement survive due diligence and any subsequent renegotiation, which is the first point at which the claims meet records. On air, founders' numerical claims are difficult to verify, and conflicting interests may

limit how much information they convey (Crawford and Sobel, 1982). Yet the decision to disclose a figure, or to withhold it, may itself be informative (Grossman, 1981; Milgrom, 1981). Early-stage financing is therefore not one decision repeated three times: each stage involves a different question, and the information available to evaluate it changes.

The paper makes three contributions. First, we build a new dataset of 642 founder–investor negotiations, transcribed and checked word by word, and linked to what became of each venture, including whether an agreed deal was ever completed, which no database records. Second, we shift the unit of analysis in the literature on entrepreneurial finance from a single funding outcome to the sequence of decisions that produces it. Existing work has studied screening, evaluation, contracting and due diligence largely on separate samples or margins (Kaplan and Strömberg, 2001, 2004; Bernstein et al., 2017; Gompers et al., 2020; Fitzgerald and Ray, 2025; González-Uribe et al., 2026). We observe several of these decisions for the same ventures and show that the information associated with success changes across them. A conclusion about what drives early-stage financing therefore depends on which decision is being studied. Third, we provide a new way to measure dimensions of the qualitative information on which early-stage investors are known to rely. Conventional textual analysis quantifies features of language; LLMs allow us to construct assessments from the content of a pitch in context. In our data, judgment predicts who draws competing offers, what terms a founder gets, and which ventures survive. Once judgment is included in the regression, conventional textual analysis no longer predicts who draws offers or which ventures survive, and retains only a weak association with the terms. Our contribution is not that an algorithm predicts investment but that contextual judgment provides a systematic measure of a previously difficult-to-measure dimension of soft information. This measure allows us to identify the stages of the investment process at which that judgment is informative and those at which it is not.

This paper contributes to three strands of literature. First, we contribute to the literature on how early-stage investors make decisions. When asked directly, investors report that they bet on the team first and on the business model, the product and the market next, and rank valuation behind all of these factors (Gompers et al., 2020). Their own investment analyses record the risks they weigh and tie them to the terms they write (Kaplan and Strömberg, 2004); angels screen ventures by looking for fatal flaws rather than by scoring strengths (Maxwell et al., 2011); and investors' network connections are associated with the performance of their investments (Hochberg et al., 2007). Other work isolates particular signals. In a randomized experiment on written profiles, information about the team raises investor interest (Bernstein et al., 2017). In pitch videos, delivery predicts who gets funded (Hu and Ma, 2025), and first impressions predict investment offers, although the impressions that track the venture's later performance are not the same ones that track the offer (Huang et al., 2023). The questions investors ask and the way founders answer them are related to funding outcomes (Kanze et al., 2018), and the narratives firms tell are associated with the resources they raise (Martens et al., 2007). These papers identify what investors respond to, but they leave open whether the information behind one decision is also the information behind the next. Most of the

evidence links a particular signal to a particular margin; it does not compare, for the same venture, how the information associated with the decision changes from *interest* to *handshake* to *completion*. We confirm that soft information is central to the initial screening, and extend this literature by showing that the forward financial claims that matter little for drawing interest are what separate the agreements that close from the ones that do not.

Second, we contribute to the literature on machine learning in finance. A long tradition extracts economic signals from text: the tone and readability of disclosures predict outcomes (Loughran and McDonald, 2011; Li, 2008), pessimistic media text predicts short-run price pressure followed by reversal (Tetlock, 2007), and this approach has been reviewed in surveys of text as data (Gentzkow et al., 2019; Loughran and McDonald, 2016). Machine learning has since moved from measuring the text to measuring behavior within it, such as non-answers to analysts' questions (Barth et al., 2023), and then into financial decisions themselves: credit scoring from digital footprints and alternative data (Berg et al., 2020; Jagtiani and Lemieux, 2019), with consequences for credit access and pricing across borrower groups (Fuster et al., 2022; Bartlett et al., 2022), and the identification of skill among fund managers (Kaniel et al., 2023). Machines have also been turned on the decision-makers: they complement rather than replace analysts (Cao et al., 2024), and much of what mutual-fund managers trade can be predicted from what they traded before (Cohen et al., 2026). Large language models now read financial text directly, analyzing financial statements (Kim et al., 2024) and forecasting returns (Lopez-Lira and Tang, 2026; Chen et al., 2022). Closest to our setting is the work of Lyonnet and Stern (2024), who compare machine-learning predictions of venture performance with the ventures that venture capitalists actually fund. While they evaluate investors' choices against a machine benchmark, we examine at which stages of the investment process machine judgment is informative and at which it loses its predictive power.

Third, we contribute to the literature that uses artificial intelligence as a research instrument rather than studying it as a subject. Machine learning can be used to find patterns in high-dimensional data and turn them into interpretable hypotheses, rather than to score a list drawn from the researcher's priors (Ludwig and Mullainathan, 2024). Following this approach, we let the models generate the judgment variables themselves. Language models have been used as simulated economic agents (Horton et al., 2023), as interviewers in qualitative research (Chopra and Haaland, 2023), and in the field on both sides of a job interview: conducting it, with human recruiters evaluating what it elicits (Jabarian and Henkel, 2026), and assessing it against professional recruiters (Avery et al., 2026). Using LLMs in this way raises its own measurement question, because repeated requests to the same model need not return the same output (Coqueret et al., 2026). We therefore treat the judgment scores as noisy measurements rather than fixed labels, and Section 6.9 reports which parts of our measure are reproduced and which are not.

The paper proceeds as follows. Section 2 describes the *Shark Tank* setting, the sample and three investment decisions. Section 3 sets out the conceptual framework and the four kinds of information. Section 4 describes how each is measured and how the four indices are built. Section 5 states the empirical strategy. Section 6 reports the results: how each kind of information predicts the three

decisions, what drives each index, the terms of the agreement, what the ventures became, train–test prediction, and what holds under alternative choices. Section 7 concludes.

2 Data

Empirically, studying early-stage investment decisions is difficult, as founder-investor conversations typically take place behind closed doors, in private meetings, pitch events, and informal networking environments, leaving no public record of what was said, how founders presented themselves, or how the two sides interacted. Existing databases, such as PitchBook and Crunchbase, record financing outcomes, but the interactions that shaped those outcomes remain unobserved. To address this limitation, we construct a novel dataset of founder–investor interactions using the U.S. business reality television show *Shark Tank*. Although the show is subject to entertainment-driven editing, it provides one of the few publicly available settings in which fundraising interactions between entrepreneurs and investors can be observed, including entrepreneurs presenting business opportunities, investors asking follow-up questions, and the resulting investment decisions. Additionally, we track each venture after it leaves the set, from the open web, and as supplementary evidence, from Crunchbase.

Shark Tank has aired on ABC since its premiere on August 9, 2009, and, as of July 2026, has run for seventeen seasons (American Broadcasting Company, 2026). Across those seasons, the show has aired 359 episodes, according to the published episode listings. The show follows a straightforward and consistent format. In each episode, entrepreneurs pitch their businesses to a panel of five investors, known as the "Sharks", to seek funding. The panel is drawn from a small group of regulars, with a guest investor taking a regular's seat in roughly half of the pitches. After introducing themselves and their business, the entrepreneurs deliver a pitch explaining their product or service, business model, market opportunity, and funding request. Many also show their products through live demonstrations or prototypes to showcase their functionality and appeal. Following the pitch, the "Sharks" engage with the entrepreneurs in an extensive question-and-answer session, probing the venture's financial performance, sales, margins, competition, valuation, growth strategy, and the founders' experience. Based on the information revealed during this discussion, any interested "Sharks" from the panel may make an investment offer. Multiple "Sharks" may compete with individual offers or join together in a single deal, while entrepreneurs are free to negotiate, make a counteroffer, accept an offer, or leave the show without an investment if no agreement can be reached.

The unique setting of *Shark Tank* has been used in prior empirical work. Lavanchy et al. (2022) construct a hand-collected dataset of pitches to examine real-time valuation negotiations, finding that smaller equity asks attract more offers and that bargaining centers on the equity share. Jetter and Stockley (2023) use manually coded founder-investor interactions to study gender-match effects in investment decisions, and Huang et al. (2023) analyze video stills from the show to investigate the role of first impressions in angel investing. Beyond academia, *Shark Tank* is deemed to provide

a useful lens through which to observe evolving entrepreneurial activity and consumer demand, as the products pitched and funded on the show reflect broader changes in the U.S. economy over time (The New York Times, 2024). The show has also won five *Emmy Awards*, including the 2024 award for *Outstanding Structured Reality Program*, from 28 nominations (American Broadcasting Company, 2026).

2.1 Building the dataset

2.1.1 On-air transcripts

We transcribed and standardized all available 642 pitches from 229 episodes on YouTube between February and April 2025, with further transcription in April 2026, and human research assistants checked every transcript word by word against the recording.[1] We structured each transcript as a dialogue between the founder(s) and the investors, preserving the sequence of the interaction from opening pitch through negotiation and final comments. We tagged every line with the aspect of the fundraising interaction it addresses: product and technology, founder and team, financials and valuation, commercial strategy, market and competition, operations and execution, investor value added, negotiation, final comments, and non-substantive talk (Appendix C.1.1 provides an example of a full transcript, shown before cleaning). From these transcripts we construct four sets of variables: the realized financial figures the founder states, the founder's forward financial claims, counted features of how the founder speaks, and a large language model's judgments of venture quality (described in Section 4).

Our variables must capture what an investor could know at the moment of the decision, and must not encode the decision itself to prevent data leakage. The risk is specific to our setting: one transcript carries both the pitch characteristics we measure and, in its raw form, the outcome we want to predict. A variable that drew on the negotiation would therefore not predict the handshake; it would restate it. Two features of the raw transcripts specifically create that risk, and a third creates a confound. First, the raw transcript records the negotiation, so it states the offer terms, the entrepreneur's acceptance or rejection, and the reactions that follow. Second, it also names the founder(s), the venture, the product and the investors, which could let a language model identify the pitch and draw on what it already knows about the outcome rather than read what was said. Third, English language marks gender in pronouns, honorifics, and kinship terms, which would let a language model condition its assessment on founder gender.[2] We removed all three.

We first anonymized the transcripts. We collapsed every speaker label to a generic role (`INV`, `FOUNDER`) and replaced the names within the dialogue itself: a founder's name by `FOUNDER`, an investor's name by `INV`, the venture's name by `Company`, and a separately branded product's name by `Product`. These placeholders therefore also appear inside the speech, where founders introduce

[1] All pitches were transcribed from the YouTube channel Shark Tank Global using the Upmeet AI platform; see Appendix C.1 for further details.

[2] The removal is a precaution rather than a correction for an observed gap: only-female and only-male teams attract investor interest at 50.0 and 49.8 percent ($p = 1.00$) and reach a *handshake* at 68.4 and 65.6 percent ($p = .60$).

themselves, name their venture and products or address an investor by name. For example, a pitch opens "Hi, my name is `FOUNDER`, and my company is called `Company`."

We then neutralized gendered language. We removed honorifics (Mr., Ms., Sir), replaced gendered pronouns and possessives with their plural equivalents (he/she → they, him/her → them, his/her → their), and replaced gendered kinship terms with neutral ones (mother/father → parent, brother/sister → sibling, husband/wife → spouse, son/daughter → child, boyfriend/girlfriend → partner). No gendered pronoun or honorific remains in the cleaned transcripts. We kept gendered nouns that describe a market rather than a speaker, since a phrase such as "underwear designed for women" is information about the venture.

Finally, we removed all information relating to the outcome. We dropped every investor turn, and from the founder's own speech, we dropped the lines tagged as negotiation and as final comments. These are the three places where the offer terms, the acceptance or rejection, and the post-decision reactions appear. Dropping the investor turns also removes what their questions reveal about investor engagement before any offer is made. The remaining eight categories are retained in full. The cleaned transcript therefore contains what an investor observes before deciding: the founder's own account of the venture, in the order it was delivered. For example, in the Chill Soda pitch (Season 1, Episode 9), the raw transcript carries the full negotiation: "The offer is for the $50,000, but I need to get 20% of your business," the founder's acceptance, the reactions on both sides. The cleaned transcript retains the opening pitch, the product description, the ask, the business background, and the founder's answers without the questions that prompted them. The cleaned transcript is the input to every variable we construct. Appendix C.1.1 shows the Chill Soda transcript as recorded and Appendix C.1.2 the same pitch after cleaning, allowing a side-by-side comparison.

2.1.2 Post-show outcome

Each pitch ends with an on-air decision. That decision tells us whether an investor and a founder reached an agreement, a *handshake*, but not whether the investment was completed or how the venture performed afterward. The distinction matters because an on-air *handshake* is an agreement subject to due diligence rather than a completed investment, and a substantial share of agreed deals are later renegotiated or abandoned (Canal, 2016). In our data, 39.3 percent of the agreed deals whose outcome we can verify fell through.

We therefore track each venture after its appearance on the show: whether the *handshake* was completed, whether the venture was still operating one, two, and three years later, whether it raised capital again, and whether it was acquired. No single structured database records these post-show outcomes comprehensively. Whether a televised *handshake* was honored is not recorded in any database, and a small venture that stops operating rarely announces it. Both exist only in scattered public traces: a founder's later interview, a dead storefront, a local newspaper, and a state filing. Language-model agents can now collect such records at scale, and we treat that

capability as an instrument whose reliability we document.

The post-show collection is an open-web search, run twice as a two-pass replication in September 2026.[3] Each pass split the 642 ventures into 65 batches and assigned each batch to a new language-model agent with no memory of the other batches, no access to our data, and no knowledge that a second pass existed. Each agent searched the open web venture by venture, recorded every URL it opened, and coded an outcome only from pages it had opened in that session. The full instruction is reported in Appendix C.4.

Two rules govern what survives. Within a pass, the agents recorded a value whenever they could establish it and, separately, marked whether two independent sources supported it, with independence defined over five families of sources; that marking records how deep the evidence is. Across passes, we retain an outcome only where the two passes independently returned the same reading, or where it follows from a reading they agreed on: a venture both passes record as operating at the collection date is treated as having survived every horizon before that date. Where they disagreed, or where only one found anything, the outcome stays unverified.

How often the two passes agree is our check on the collection's reliability. Of the 429 *handshakes*, both passes returned a *completion* outcome for 334 and agreed on 313 of them, 93.7 percent; of those, 190 closed and 123 fell through. Agreement between passes is higher on survival: the passes agreed on 615 of the 620 ventures for which both returned first-year survival (99.2 percent), on 533 of 535 in the second year (99.6 percent), and on 490 of 494 in the third (99.2 percent); of these, 98.4, 94.6 and 90.6 percent were still operating. For current operating status, they agreed on 593 of 618 (96.0 percent). Survival is observed over the window each season allows: the first year after airing for all sixteen seasons, the second for Seasons 1–15, the third for Seasons 1–14.

Completion is the hardest of these to document and the most thinly sourced: of the 313 retained outcomes, 67 rest on two independent sources in both passes and 147 on a single source in both, with the remaining 99 resting on two sources in one pass and a single source in the other. It is also where the passes disagree most often, 21 cases against five for first-year survival. We report this rather than smooth it, because a deal abandoned quietly leaves less public trace than one completed, and the direction of that asymmetry matters for how the completion results are read. Appendix A.4 reports evidence depth field by field.

We also match each venture to Crunchbase, which records rounds announced after the episode aired and acquisitions. A profile exists for 388 of the 642 ventures; within them, 138 have a post-airing round recorded and 18 were acquired. We keep this evidence separate from the collection above and treat it as supplementary.

2.2 Three decisions: Investment funnel

The data captures three distinct decisions in the investment process. An investor first decides whether a venture is worth competing for, which we define as *investor interest*; then whether to

[3] Both passes ran on `claude-opus-5`.

agree on terms with the founder, which we define as an on-air *handshake*; and finally, following due diligence, whether the agreed investment is ultimately completed, which we define as *completion*. These three decisions form an investment funnel, but each is a distinct margin: the outcome at one stage does not determine the outcome at the next.

Investor Interest. A pitch attracts *investor interest* if it receives at least two competing offers. We use multiple offers rather than a single offer to capture strong investor demand for the venture before an agreement is reached. *Investor interest* therefore measures whether a venture attracts competing investors, rather than whether it ultimately receives funding. Additionally, a single offer is a necessary condition for a *handshake*, so *investor interest* defined as at least one offer would be mechanically tied to it: no pitch without an offer reaches an agreement. At two offers, the margins stay distinct: 40.3 percent of pitches without competing offers still reach a handshake.

Handshake. A pitch reaches a *handshake* if the founder and an investor agree to a deal on air. Unlike interest, this outcome requires both parties to agree on the terms of the investment.

Completion. A *handshake* reaches a *completion* if the agreed investment closes after the show. Following the on-air agreement, investors conduct due diligence to verify the information presented during the pitch. Either party may subsequently renegotiate or withdraw. *Completion* therefore captures whether an on-air agreement ultimately results in capital being invested.

Investor interest and the *handshake* represent distinct on-air margins. A pitch may reach an agreement after receiving only one offer, which happens in 128 of the 429 handshakes, while 23 pitches that attract competing offers leave without an agreement. We therefore treat *investor interest* and the *handshake* as separate decision margins rather than as an ordered sequence. *Completion*, by contrast, is observed only conditional on a *handshake*. Hence, we examine separately what predicts *investor interest*, what predicts an on-air *handshake*, and what ultimately determines the *completion* of an agreed deal.

2.3 Descriptive statistics for ventures

Table 1 summarizes the pitch, deal, and post-show characteristics of our sample. Panel A describes founder and team characteristics. At the team level, 50.3 percent of the 642 pitches are delivered by only-male teams, 27.1 percent by only-female teams, and 22.6 percent by mixed-gender teams; 46.6 percent of pitches are delivered by a solo founder. Teams are small, with 1.6 presenters on average. At the individual level, the sample includes 1,025 founders across these 642 pitches, of whom 37.2 percent are female.

Panel B reports deal terms. Founders ask for $303 thousand on average in exchange for 10.9 percent equity, implying an average asked valuation of $4.6 million. Among the pitches that reach a *handshake*, the average amount is slightly higher, at $317 thousand, but for 21.4 percent equity, and the average agreed valuation is $2.7 million. Founders who reach a *handshake* thus give up nearly twice the equity they initially offered (21.4 against 10.9 percent), and among the 423 *handshakes* with a recorded valuation, the agreed valuation averages $1.7 million below the valuation implied

by the ask.

Panel C reports the five largest industry categories by pitch count. Manufacturing accounts for the largest share of pitches, 46.7 percent, followed by Retail Trade, 12.0 percent; Agriculture, Forestry, and Fishing, 7.9 percent; and Other Services and Arts, Entertainment, and Recreation, with 5.3 and 5.1 percent, respectively. The remaining industries, each individually small, are pooled together and account for 22.9 percent of the sample.

Panel D reports the composition of the investor panel. The average panel includes five investors; four pitches face six. A guest investor sits on the panel for 45.6 percent of pitches, in place of a regular rather than in addition to one. At least one female investor is present in every pitch, and the average panel includes 1.6 female investors. Among the 428 *handshakes* for which the agreeing investors are recorded, 53.2 percent are agreed exclusively with male investors, 28.5 percent exclusively with female investors, and 18.2 percent jointly with male and female investors.

Panel E reports the on-air investment outcomes. Pitches receive 1.6 offers on average, and 66.8 percent reach an on-air agreement. We define *investor interest* as receiving at least two offers, which occurs for 50.5 percent of pitches. Attracting *investor interest* and reaching a *handshake* are related but distinct. Of the 429 pitches that reach a *handshake*, 301 receive at least two offers, while 128 receive only a single offer. At the same time, multiple offers do not guarantee a *handshake*: 23 pitches with two or more offers leave without an agreement. The probability of a *handshake* rises sharply with the number of offers, from 76.2 percent for pitches receiving a single offer to 88.3 percent for those receiving two and 98.6 percent for those receiving three or more. Many pitches with a single offer still reach an agreement, while some pitches attracting several offers do not.

Panel F reports post-show outcomes, all taken from the open-web collection described above. We document *completion* for 313 of the 429 on-air *handshakes*. Among these, 60.7 percent completed the investment after due diligence and 39.3 percent fell through, showing that an on-air agreement does not necessarily lead to a completed investment. We also track subsequent venture outcomes. Among the 593 ventures for which we establish current operating status, 86.0 percent are active or have been acquired. Survival is observed over the window each season allows: of the ventures with a verified outcome, 98.4 percent were still operating one year after airing (619 ventures, Seasons 1–16), 94.6 percent after two years (540, Seasons 1–15), and 90.8 percent after three (498, Seasons 1–14).[4] Beyond survival, 7.0 percent of the 587 ventures with a verified acquisition status were bought by another party, and 15.4 percent of the 371 ventures for which we can establish it raised a new investor round after airing.

Taken together, Panels E and F show that our three outcome variables, *investor interest*, *handshake*, and *completion*, capture distinct stages of the investment process. About half of all pitches attract competing offers, two-thirds reach an on-air agreement, and about three in five of the verifiable on-air agreements ultimately translate into actual investments.

[4] A venture recorded as operating at the time of collection is coded as having survived every earlier horizon within its window. A venture recorded as closed enters a horizon only if its closure date is verified; 35 closed ventures have no verified closure date and therefore enter no horizon. First-year survival is measured over all sixteen seasons.

Table 1: Pitch, On-Air, and Post-Show Characteristics

	N	Mean	SD	Min	Max
A: Founder and team characteristics					
A1: Team level					
Number of presenters (team size)	642	1.60	0.63	1	5
Solo founder	642	0.466	0.499	0	1
Team = only male	642	0.503	0.500	0	1
Team = only female	642	0.271	0.445	0	1
Team = mixed	642	0.226	0.418	0	1
A2: Individual level					
Female founder	1025	0.372	0.483	0	1
B: Deal terms					
Asked amount (thousand)	642	303	281	10	2500
Handshake amount (thousand)	429	317	304	10	3000
Equity offer (%)	642	10.9	6.5	1.0	50.0
Handshake equity (%)	428	21.4	14.3	0.0	100.0
Asked valuation (million)	642	4.55	6.19	0.04	50.00
Handshake valuation (million)	423	2.67	4.23	0.05	36.00
C: Industry (five largest by pitch count)					
Manufacturing	642	0.467	0.499	0	1
Retail trade	642	0.120	0.325	0	1
Agriculture, forestry, and fishing	642	0.079	0.271	0	1
Other services	642	0.053	0.224	0	1
Arts, entertainment, and recreation	642	0.051	0.221	0	1
Other industries	642	0.229	0.420	0	1
D: Investor panel					
Number of investors	642	5.01	0.08	5	6
Number of female investors	642	1.57	0.60	1	3
Guest investor present	642	0.456	0.498	0	1
E: On-show outcomes					
Number of offers received	642	1.61	1.28	0	5
Investor interest ($\geq$2 offers)	642	0.505	0.500	0	1
Handshake (1 = reached an agreement)	642	0.668	0.471	0	1
F: Post-show outcomes					
F1: Deal completion					
Completion outcome documented	429	0.730	0.445	0	1
Deal completed (of the documented)	313	0.607	0.489	0	1
F2: Venture survival, acquisition and follow-on					
Operating (active or acquired), current	593	0.860	0.347	0	1
Alive at year 1 (Seasons 1–16)	619	0.984	0.126	0	1
Alive at year 2 (Seasons 1–15)	540	0.946	0.226	0	1
Alive at year 3 (Seasons 1–14)	498	0.908	0.290	0	1
Ever acquired	587	0.070	0.255	0	1
Raised a new investor round	371	0.154	0.361	0	1

Notes. Sample: 642 pitches, *Shark Tank* Seasons 1–16 (2009–2025).
Panel B: an agreed valuation is the agreed amount divided by the agreed equity share, so it exists only where both are recorded and the share is not zero. Of the 429 handshakes, five are royalty agreements that involve no equity and one records no equity share at all, which is why the agreed equity rests on 428 handshakes and the agreed valuation on 423.
Panel E: *Handshake* equals one if the entrepreneur left with an agreement on air.
Panel F reports post-show outcomes; every one of them from the open-web collection, retained only where the two independent passes returned the same reading.
F1: of the 429 on-air *handshakes*, a *completion* outcome is independently documented for 313; among those, 60.7% were completed after due diligence and 39.3% fell through. The remaining 116 are undocumented.
F2: Each survival horizon uses the seasons for which that many years have elapsed by the collection date: year 1 Seasons 1–16, year 2 Seasons 1–15, year 3 Seasons 1–14. A venture operating at the collection date is coded as having survived every earlier horizon within its window. *Ever acquired* is the venture having been bought by another party; the collection records no public offerings, so these are trade sales. *Raised a new investor round* is a named investor round raised after the episode aired and does not count a closed on-air deal, so it measures capital raised beyond the investment agreed on the show. Denominators (N) give the number of pitches with a verifiable value.

2.4 Descriptive statistics for investors

The investment decisions in our setting are made by active investors. Across the sample, 38 distinct investors appear on the panel, eight of them regular investors and 30 guests. The regular panel includes Mark Cuban, Lori Greiner, Kevin O'Leary, Barbara Corcoran, Robert Herjavec, Daymond John, and, in the first two seasons, Kevin Harrington. Daniel Lubetzky appears as a guest from Season 11 on and joins the panel as a regular in Season 16, the last season of our sample. Several built and sold a business before joining the show, and several invest professionally through their own funds. Their industry experience ranges from consumer products and retail to technology, real estate and cybersecurity.

Table 2 summarizes the business and investment backgrounds of the regular investors. The companies they founded include Broadcast.com, FUBU, The Corcoran Group and BRAK Systems. Their investment activity also extends beyond *Shark Tank*. Crunchbase records outside investments for 33 of the 38 investors, 842 in total, although these figures provide only a lower bound because investments without an institutional funding round are less likely to be recorded on Crunchbase. Several invest through named vehicles rather than personally, among them O'Leary Ventures, Forefront Venture Partners, Camino Partners, CAVU Consumer Partners, RSE Ventures and Sound Ventures. The guest investors have similarly substantial operating or investment experience and include founders and executives from companies such as Spanx, Stitch Fix, TOMS, Ring, KIND Snacks, DoorDash, Nextdoor and Raising Cane's. Their detailed backgrounds are reported in Appendix Table A.1.

Taken together, the investor backgrounds and their activity outside the show indicate that the investment decisions observed in our setting are made by individuals with substantial entrepreneurial and investment experience. The capital at stake is also their own: a disclaimer at the start of each episode states that the money committed by the investors is theirs rather than the network's.

2.5 Data limitations

Even though *Shark Tank* offers a rare window into founder-investor interactions, it comes with limitations. It is a television program rather than a random sample of entrepreneurial ventures. Producers select pitches in part for their entertainment value, so the set of aired pitches is subject to editorial selection.

Our sample adds a second layer of selection. We observe 642 of the 1,442 pitches that aired in Seasons 1–16 (44.5 percent), those whose episodes are available on the show's YouTube channel. Coverage is uneven across seasons and rises over the show's run, from 9 percent of pitches in Season 1 to 88 percent in Season 16, and from 20 percent across the first eight seasons to 68 percent across the last eight. *Handshake* rates rose over the same period: in the published listings, the share of aired pitches reaching a deal goes from 42 percent in Season 1 to 73 percent in Season 16, and in our own sample the *handshake* rate is 57 percent over Seasons 1 to 8 against 70 percent over

Table 2: Background and Investment Activity of Regular Investors

Investor	Entrepreneurial background	Business outcome	Crunchbase activity		Shark Tank
			Investments	Exits	Sample deals
Mark Cuban	Founded Broadcast.com; founded Cost Plus Drugs (2022)	Broadcast.com to Yahoo!, $5.7bn (1999)	282	49	128
Lori Greiner	Inventor and QVC retailer; 120 patents in consumer housewares	Founded For Your Ease Only (1996); still president	21	1	127
Kevin O'Leary	Founded SoftKey/The Learning Company	Sold to Mattel, $4.2bn (1999)	19	4	71
Barbara Corcoran	Founded The Corcoran Group (real estate business)	Sold to NRT, $66m (2001)	11	1	57
Daymond John	Founded FUBU (clothing company for young men)	FUBU, reported $6bn in lifetime sales	10	2	51
Robert Herjavec	Founded BRAK Systems (security software) and the Herjavec Group (Cyderes)	BRAK sold to AT&T Canada, $30.2m (2000); Herjavec Group merged into Cyderes (2022)	15	2	47
Kevin Harrington	Infomercial pioneer; founded Quantum International	Quantum, reported $500m in annual sales; NYSE-listed	3	–	2
Daniel Lubetzky	Founded KIND Snacks	KIND to Mars, $5bn valuation (2020)	12	2	1

Notes. Investors are ordered by the number of agreements reached. *Investments* and *Exits* are from Crunchbase, 28 September 2026; an exit is a portfolio company acquired or taken public. *Sample deals* counts the on-air agreements each investor took part in while seated as a regular, so a joint agreement is counted once per participant: the column gives 484 pairs across 388 of the 429 agreements, the remaining 41 being agreements reached by guest investors alone. Daniel Lubetzky sat as a regular only from Season 16, so his column entry is 1; the 17 further agreements he reached as a guest investor in earlier seasons are counted among the guest agreements and are reported in Appendix Table A.1. Backgrounds and business outcomes are from company filings, press releases, trade press and published biographies, September 2026.

Seasons 9 to 16. Our sample is therefore weighted towards the seasons in which deals are most common, and the pitches we observe reach a *handshake* more often than those we do not, 66.8 against 56.9 percent, though the latter rests on published deal counts rather than on our own coding of a *handshake*. Whether that gap reflects the seasons we observe or the pitches within them we cannot say, because outside our sample we have only published season totals rather than a pitch-level record. Appendix A.2 reports what aired against what we observe, season by season.

The televised episode is also a heavily edited version of a substantially longer negotiation. Executive producer Clay Newbill reports that the actual pitches run forty-five minutes to an hour and are edited down to roughly seven minutes on air (Wallace, 2025); Barbara Corcoran has described days of ten to twelve pitches, each lasting between forty-five minutes and an hour and a half (CNBC Make It, 2018). The broadcast may therefore omit parts of the financial discussion, the valuation negotiation, and the questioning that informed investors' decisions.

Our measures are moreover built from transcripts, which lose tone of voice, body language, and the product demonstrations that investors observe in the room. Our estimates therefore capture the information in what founders say, not in how they appear or what they show.

An on-air agreement does not necessarily correspond to a completed investment. After due diligence, either party may renegotiate or withdraw. Canal (2016) contacted 237 of the 319 businesses that accepted a deal on air in the first seven seasons: 73 percent did not receive the deal seen on television, and about 43 percent reported that the deal never came to fruition. We also find that among the handshakes whose outcome we can verify, 39.3 percent fell through. Post-show outcomes are, however, observed only where public sources document them. We verify completion for 313 of the 429 handshakes. Documented and undocumented handshakes do not differ in the ask, the equity offered, the implied valuation, revenue, the number of offers, or industry; the share of female founders differs weakly (Mann-Whitney test $p = .020$), and they differ in season ($p < .001$), since later seasons have had less time to be documented. Balance on these characteristics does not rule out selection on the outcome itself. A deal that is abandoned quietly leaves no public record, while a completed one is often announced by the venture or the investor, so abandoned deals are the harder of the two to document. We therefore read the completion stage against the two stages that precede it, which are observed for all 642 pitches, rather than on its own. Closure dates are also harder to verify than current operating status, so survival at fixed horizons rests on fewer ventures than current status does.

Shark Tank is also a particular corner of early-stage finance. Nearly half the ventures we observe are consumer manufacturers, the median ask is $250,000, and the investors decide in a single televised meeting rather than over weeks of private diligence. The setting is therefore closer to angel investment in physical consumer products than to institutional venture capital. Whether investors weigh soft information in the same way when the sums are larger, the diligence is longer, and the entrepreneur is met in private is a question this setting cannot answer.

Finally, although the primary objective of participants is to obtain funding, some entrepreneurs also appear to value the substantial publicity generated by the show, so exposure may complement capital raising for a subset of founders.

Despite these limitations, *Shark Tank* remains one of the few publicly observable settings that records the complete sequence of entrepreneurial pitching, investor questioning, negotiation, and investment decisions.

3 Conceptual Framework

Founders know more about their ventures than investors do, and this information asymmetry is particularly severe for young, early-stage firms (Gompers, 1995). Investors respond by screening before they commit capital (Kaplan and Strömberg, 2001). Screening is not a single decision. Venture capitalists move an opportunity through successive stages, from a first screen to a meeting with management, a partner review, due diligence, a term sheet, and closing, and they invest in

roughly one of every hundred opportunities they consider (Gompers et al., 2020). Nor is it passive. At each stage investors question the founder and probe the claims made, so part of what they learn is information they elicit themselves. Each stage therefore resolves part of the adverse-selection problem, and the natural question is what kind of information it is resolved with.

Shark Tank compresses this sequence into three observable steps. Investors first decide whether a venture is worth competing for (*investor interest*), the two sides then agree on terms on air (the *handshake*), and the agreement is completed or abandoned after due diligence (*completion*). The framework classifies the information in a pitch into four kinds and measures each of them identically at every stage, so that the stages can be compared by which kind of information predicts them.

3.1 Hard and soft information

We build on Liberti and Petersen (2019). Hard information is quantitative, and its meaning does not depend on the context in which it was collected: the receiver of the data knows everything relevant that the sender knows. Soft information is information whose context, and whose collector, are part of the information itself. It "includes opinions, ideas, rumors, economic projections, statements of management's future plans, and market commentary" (Liberti and Petersen, 2019, p. 4). The distinction coincides with verifiability in the contracting sense: a signal is verifiable when every party interprets it the same way (Liberti and Petersen, 2019). We operationalize it accordingly. Information is hard if its value is fixed by the transcript, so that any reader working from the same words would record the same figure. A stated revenue figure is hard in this sense. Whether a founder understands their customers, or whether an ask is reasonable for this particular business, is a conclusion a reader draws from the pitch as a whole, and two readers can draw different conclusions.

Three points from Liberti and Petersen (2019) shape our measurement. First, a number is not automatically hard. A score of how trustworthy a founder is remains soft unless every reader would assign the same score, and, conversely, a projection is soft even though it is stated as a figure, because its content is a claim about what has not happened rather than a record of what has. Second, hard is not the same as true. Founders state their own figures, which investors cannot check on the spot and can verify only in due diligence after the handshake. Third, the distinction is a continuum rather than a dichotomy, and soft information can be partly *hardened*, that is, converted into a numerical index, at the cost of some of the context that gave it meaning.

Early-stage equity is the setting in which this distinction bites hardest. Equity is among the most opaque asset classes: young firms have no track record, so little hard information exists, and investors rely on soft information gathered through personal contact (Liberti and Petersen, 2019). Soft information is difficult to pass on without the person who collected it, so the collector and the decision maker are typically the same person (Stein, 2002). In *Shark Tank*, they are: the investors who question the founder are the investors who decide.

The distinction also determines how credible a claim is. A founder who states a figure that due

diligence can later check has little reason to misstate it, and a founder who withholds such a figure signals that it is unfavorable (Grossman, 1981; Milgrom, 1981). A claim that cannot be checked, such as a description of the market or a projection of future revenue, costs nothing to exaggerate and is therefore only partly informative (Crawford and Sobel, 1982).

3.2 Four kinds of information

Hard and soft information alone do not separate what a pitch contains, for two reasons. Numbers split according to whether they record something realized or assert something unrealized. And soft content can be measured in two ways: hardened by counting, or read in context. We therefore distinguish four kinds of information, which lie at different points of the continuum.

Hard finance (H_{finance}) consists of the realized figures the founder states: revenue to date, units sold, prices, unit costs, margins, marketing and fixed costs, funding raised, debt and profitability. Two readers working from the same transcript record the same values, and each figure describes something that has already happened and can in principle be checked against records. This is hard information in the full sense of Liberti and Petersen (2019).

Soft finance (S_{finance}) consists of the founder's own forward claims, stated as numbers: expected revenue, market size in revenue and in customers, the payback period, and the ask itself, that is, the capital requested, the equity offered, and the valuation the two jointly assert. These are as reproducible as the realized figures, and they are not hard. Liberti and Petersen (2019) count economic projections and statements of management's future plans as soft information, and the reason is not that the number is imprecise but that its content cannot be separated from who produced it and under what incentive. The ask enters here because a request for a given sum at a given equity share asserts a valuation, and a valuation is a claim about what the venture will be worth.

Language (S_{language}) is soft information hardened by counting. Word counts, dictionary categories and sentiment scores are reproducible by construction: anyone running the same code on the same transcript obtains the same values. They achieve this by discarding context. A count records that a founder used the word "confident" but not what they were confident about. Liberti and Petersen (2019) describe exactly this trade-off for textual analysis in finance: text condensed into numerical indexes captures only part of what a human reader would take from the original. Language is therefore hard in how it is measured and soft in what it measures, which is why it sits between the stated figures and judgment rather than at either end.

Judgment (S_{judgment}) is soft information read in context: assessments of product-market fit, execution risk, or the credibility of the plan that a reader forms from the pitch as a whole. This is hardening of a different kind, and one that has only recently become possible at scale. Where a dictionary count strips the context, a language model retains it: it can relate an ask to the rest of the pitch rather than to a fixed rule. Judgment therefore arrives as numbers, but it is not hard, because different readers assign different numbers. Two language models scoring the same variable

on the same transcripts agree at a median rank correlation of 0.66 across the variables that enter the index. A stated figure such as the ask leaves no such room: it is read off the transcript rather than inferred from it.

These four kinds sit at different points on the hard-soft continuum: hard finance at one end, soft finance next because its figures are precise but unverifiable, then language, which is reproducible only given a choice of dictionary, and judgment at the other end, where two readers disagree.

4 Measurement

Every pitch is represented by the four information sets of Section 3 and by a set of structural controls. All four sets were built from the same cleaned transcript. The cleaned transcript excludes the negotiation and the closing remarks, so no set contains the offers, the counteroffers, or the outcome. Founder, investor and company names were replaced by placeholders. We describe each set in turn and then the construction of one index per set.

4.1 Hard finance (H_{finance}) and soft finance (S_{finance})

Both finance indices come from the same extraction. We defined 30 variables in advance, of which 28 are stated often enough to enter the analysis together with a codebook that specifies each variable's definition and extraction rule. Claude extracted the variables from every transcript using this codebook, and we then checked each extracted value by hand against the transcript and corrected it where necessary. Every figure is either stated explicitly by the founder or derived from stated figures by a prespecified rule in the codebook, so that any reader of the transcript would record the same value. What separates hard finance from soft finance is not how they were measured but what they refer to.

Hard finance holds the 21 realized figures: current revenue, past revenue growth, units sold to date, the gross margin as stated by the founder, the gross margin computed from stated unit prices and costs, profitability, unit price and unit cost, customer acquisition cost, marketing spend, whether any marketing spend is incurred, fixed monthly costs, development expenses, debt and whether any is held, personal liability, external funding raised, prior investment, bootstrapping, capital invested by the founders, and the equity share held by non-founders. Each describes something the venture has already done.

Soft finance holds the founder's forward claims, in two groups, the projections and the ask. The projections are expected revenue, market size in customers and in revenue, and the payback period. The ask is the capital requested, the equity offered, and the valuation the two jointly imply. Two further forward-looking variables, the market share the founder expects to capture and the months of runway, are stated by too few founders to enter an index and are reported but not used. These figures are recorded as precisely as the realized ones; they are soft because their content is a claim about what has not happened, which Liberti and Petersen (2019) classify as soft information.

A variable is left empty when the founder does not state it, and most are empty often: 19 of the 28 are missing in more than 30 percent of the training pitches, while the ask, the equity offered and the implied valuation are stated in every pitch. Each variable missing in more than 30 percent of the training pitches receives a missingness dummy. The 28 variables therefore give 47 columns, 36 for hard finance and 11 for soft finance, including 19 of them missingness dummies. Non-binary variables are winsorized at the 1st and 99th percentiles of the training sample. Appendix C.2.1 lists every variable with its definition and extraction rule.

4.2 Language (S_{language})

S_{language} records how the founder speaks. Its 192 variables are computed from the founder's statements in the cleaned transcript by fixed dictionaries and parsers, with no language model involved, and anyone running the same code obtains the same values. This follows the text-as-data tradition in finance, which extracts information from how words are used (Tetlock, 2007; Loughran and McDonald, 2011), and which Liberti and Petersen (2019) describe as the hardening of soft information: the text is condensed into numerical indexes that are reproducible but retain only part of what a reader would take from it. We use established dictionaries.

The instruments use limited local context – a negation within three tokens reverses a dictionary hit, VADER applies valence and punctuation rules, readability depends on sentence boundaries – but none of them reads a claim against the venture it is about. That is the distinction this index is built to draw against S_{judgment}: language is measured mechanically from the words, judgment is read from what the words are about.

Tone and emotion (73 variables). General sentiment is measured with VADER (Valence Aware Dictionary for Sentiment Reasoning) (Hutto and Gilbert, 2014), a rule-based valence model built and validated for social media text, and discrete emotions with three NRC lexicons: the NRC Emotion Lexicon (Mohammad and Turney, 2013) for binary emotion flags, the NRC Emotion Intensity Lexicon (Mohammad, 2018) for graded intensity, and the NRC Valence-Arousal-Dominance Lexicon (Mohammad, 2025), together with Empath (Fast et al., 2016). Finance-specific tone uses the Loughran-McDonald dictionary (Loughran and McDonald, 2011), which was constructed because general-purpose word lists misclassify ordinary financial vocabulary, and which has become the standard in accounting and finance. Linguistic Inquiry and Word Count (LIWC-22) (Boyd et al., 2022) contributes its emotion categories, which count words marked as positive or negative affect.

Warmth and competence (27). The two dimensions on which people are known to evaluate others (Fiske et al., 2007) are measured with the SADCAT (Semi-Automated Dictionary Creation for Analyzing Text) dictionaries (Nicolas et al., 2021), which were validated against human ratings for exactly this purpose, together with LIWC's social and prosocial categories. Gender-typed language uses word stems in the spirit of the Bem Sex-Role Inventory (Bem, 1974).

Certainty and hedging (21). Hedges, modal verbs and expressions of confidence, from the Loughran-McDonald uncertainty, weak-modal and strong-modal lists and LIWC's tentativeness

and certitude categories. This is where a founder's conviction, or its absence, shows up in wording.

Entrepreneurial orientation (19). The entrepreneurial orientation word lists developed by Short et al. (2010) capture autonomy, competitive aggressiveness, innovativeness, proactiveness, and risk-taking in firm text, supplemented with a mentoring category.

Fluency and complexity (17). Readability, sentence and word length, type-token ratio, filler words and disfluencies: how easy the pitch is to follow, independent of what it says.

Concreteness and numeracy (10). Counts of numbers, dollar amounts and percentages, together with LIWC's quantity, space and time categories.

Narrative and reasoning (10), *self-presentation and address* (10), and *substance and time reference* (5) cover causal and insight language, pronoun use and how the founder positions themselves relative to the investors, and whether the pitch speaks of the past, the present or the future.

These nine families are an organizing and weighting device, not nine validated variables. Warmth and competence, for instance, holds the SADCAT, Eberhardt and Bem lists alongside six LIWC categories, which are not all measurements of one psychological object, and self-presentation holds LIWC's gender categories, which record a venture's market as readily as a founder's self-description. We therefore read a family coefficient as a statement about that group of dictionaries and not about the trait its name borrows. The index is built in two steps, within family and then across the nine, by the rule in Subsection 4.5. Appendix C.2.2 lists every variable with its source instrument.

4.3 Judgment (S_{judgment})

S_{judgment} contains assessments of the venture that are not stated in the transcript and must be inferred from the pitch as a whole: whether the product fits a market, whether the pricing is credible, whether the business model scales, and whether the team can execute. We use two large language models, Claude (claude-sonnet-4-6, Anthropic) and GPT-5.4 (OpenAI), and we follow the methodological literature that uses algorithms to propose interpretable measures from high-dimensional data (Ludwig and Mullainathan, 2024). The models proposed the variables themselves.

Construction proceeded in two stages. In the *variable generation* stage, each model read the 449 training transcripts in random batches of 50 and proposed economically meaningful variables that plausibly drive early-stage investment decisions, together with a codebook giving each variable's type (binary, ordinal from 0 to 4, or numeric), definition, and scoring rule. Only training pitches were used, so no test pitch influenced what is measured. The generation prompt also asked each model to judge whether the pitch ended in a deal. Because the models read only the cleaned and anonymized transcript, which excludes the negotiation and the closing remarks, this judgment was the model's own assessment and not an observed outcome.

The two models proposed 220 variables, 72 from Claude and 148 from GPT. Twenty-five variables were proposed by both, which gives 195 distinct variables. Of these, 48 restate a figure that is hand-coded from the same transcript and already enters the finance indices, and keeping them would put the same fact in two indices. Thirty-nine appear on a list of duplicated financials. Eight

more reproduce a hand-coded figure under another name, and we identified them by their rank correlation with that figure: whether the venture is bootstrapped, its debt level and whether it is debt-free, its payback period, two measures of dilution, its customer acquisition cost and its debt load, which correlate with their hand-coded counterparts between 0.76 and 1.00. The last is the gross margin, which both models proposed and which a founder states outright. Removing all 48 leaves 147 variables.

In the *scoring* stage, each model scored every transcript, in the training and the test samples, against its fixed codebook, at temperature 0.7 for Claude and with a fixed seed for GPT.[5] At both stages the models read only the cleaned and anonymized transcript and had no access to outcomes.

The 147 variables give 163 scored columns, because the 16 that both models proposed are scored twice, once by each model, and the two scores are averaged before the index is built. Counts in the paper are of variables, not of columns, except where stated. The 131 proposed by one model are scored once. The 16 proposed by both are scored by each model separately, so each of those variables has two columns, the same definition and scoring rule, applied independently by two models, which can and do return different values. Two columns are dropped in preprocessing because they have no variation among the training pitches on which the preprocessing is fitted: GPT's cash conversion, which is zero for every pitch that it scores, and its stockout measure, which is zero for every training pitch and non-zero for two test pitches. That leaves 145 variables and 161 columns. When the index is built, the two columns of a variable that both models proposed are averaged into a single score, whether or not the two models agreed, so the index rests on 145 distinct variables and no variable enters it twice.

How far the two models agree is a property of the measure worth stating, because it bounds what this index can be. On the 16 variables that both models scored, the median rank correlation between their scores is 0.66 and the median Cohen's κ is 0.44. A hand-coded figure such as the ask reaches 0.9989 between two independent extractions of the same transcript. Judgment is therefore a measurement with real disagreement in it, which is what reading an unverifiable assessment from a pitch should produce, and it is why we average the two models rather than choose between them.

The 145 variables are also grouped into nine families for a decomposition exercise: business model and economics, traction and customers, capital and deal, market and competition, operations and risk, go-to-market and distribution, product and defensibility, team and founders, and mission. We assigned each variable to the family its definition describes, and the nine partition the set. This is an organizing device for reading the index and nothing more. The families were not validated as constructs, no test establishes that they are separate dimensions, and a coefficient on one of them is a statement about that group of variables rather than about the trait whose name it borrows. The index itself is built on the 145 variables directly and does not use the grouping.

[5] Both settings leave the scoring stochastic, so the same codebook applied to the same transcript need not return the same value twice. Section 6.9 quantifies this directly: rescoring moves 4.7 percent of Claude's scores and 8.9 percent of GPT's, and we treat the scores as noisy measurements rather than fixed labels throughout.

4.4 Structural controls

The control set describes the context of the pitch rather than the venture. Eleven columns enter, in four groups: two for when the episode aired, the season and the position within it; three for who pitched, the number of presenters and two indicators for the composition of the founding team; one for whether a guest investor was seated; and five for what the venture sells, its five largest product categories. We hold the set to that size deliberately, so that the three decisions are estimated under the same conditioning variables and the completion stage, where 123 of the 313 handshakes fall in the smaller outcome class, can carry the same specification as the two stages observed on all 642 pitches. The same set carries every stage, so that the three decisions are compared under one specification rather than three. Unlike the four information sets, these variables were not taken from the transcripts: we coded them by hand from the episodes and from public episode listings, with air dates and investor panels checked against the published episode guides and the product category taken from the *Shark Tank* products website. They serve as controls and as the baseline in the prediction exercise. Founder gender enters the analysis only through these variables. Section 6.9 reports the same estimates under the full hand-coded control set, which adds the airing month and weekday, the seated investors individually, founder state, industry, and the remaining product categories. No coefficient in Table 3 changes sign under it, and one changes its significance band: soft finance at completion, entered on its own, moves from the five percent level to the ten.

4.5 From a set to an index

Each set holds too many variables to enter a regression directly, so we reduce each to one index, following the standardized-average summary index common in applied work (Anderson, 2008; Kling et al., 2007).

The rule has to settle which way each variable points, and we settle it the same way in all four sets.

A direction could in principle be imposed on the stated figures, because a figure carries its own meaning: more current revenue, more units sold, a higher gross margin and profitability describe a stronger venture, while a higher acquisition cost, more debt and higher fixed costs describe a weaker one. For the text it cannot. Nothing tells us whether longer sentences, more hedging or a particular emotion word should count as a better pitch or a worse one, and no prior literature fixes a sign for most of the 192 language variables or the 145 judgment variables. Imposing a direction on two sets and estimating it for the other two would make the four coefficients of Table 3 incomparable, since two of them would carry a direction we chose and two a direction the data chose.

We therefore let the data assign the direction in every set: each variable is standardized and then given a sign, positive if higher values go with a *handshake* and negative if they go against it. No variable is left out on the grounds that its direction is contestable, because no direction is being asserted. Section 6.9 reports the alternative for the finance indices, signing each figure by hand from its meaning, and Appendix Table C.14 sets the two side by side.

Each figure also carries an indicator for the pitches in which the founder states nothing, and these are signed by the same rule. A founder with a favorable figure has reason to state it, so silence is in principle informative about that figure rather than neutral, which is the disclosure logic of Grossman (1981) and Milgrom (1981); whether that logic applies to an edited broadcast interaction, where a figure can be missing because no investor asked, is a question we return to in Section 6.9. Fifteen of the hard finance indicators and four of the soft finance ones record such a silence; the ask, the equity offered and the implied valuation are stated by every pitch and have none.

Assigning signs from an outcome needs care, because uusing each pitch's own outcome naively would help set the sign applied to that pitch, and the index would be partly fitted to what it is meant to predict. We therefore assign the signs by cross-fitting, in the sense of Chernozhukov et al. (2018). The 642 pitches are split into five folds; the sign applied to a pitch in one fold is estimated on the other four folds only, so a pitch's own outcome never reaches its own sign. The split is redrawn 20 times, giving 100 fold estimates per variable, and the index is the average of the 20 resulting indices. Signs are not averaged into a fractional direction: within each fold a variable receives $+1$ or -1 from the sign of its correlation with the *handshake* among the training folds, and the averaging happens at the level of the finished index.

Signs estimated this way are not all equally firm, and the distribution matters more than any single summary. Of the 192 language variables, 117 hold the same sign in at least 90 of the 100 fold estimates and 50 hold it in all 100, while 17 are effectively a coin flip at under 60. For the 161 judgment columns the figures are 124, 77 and 5; for the 36 hard finance columns, 27, 15 and none; for the 11 soft finance columns, 8, 4 and none. A variable with a stable sign contributes its direction to the index; one without a stable sign is as likely to point the wrong way for a given pitch as the right way, so across 642 pitches its contributions offset and it adds close to nothing. Nothing is dropped on this basis – the unstable variables remain in the index and are simply uninformative in it – so the index is not selected on how well its parts predict.

Two details finish the construction. Where both language models scored the same variable, their scores are averaged into one, so judgment has 161 columns but 145 variables. Where a founder states no figure, the variable takes the median of the pitches that do, and the accompanying indicator records the silence. The index is then the average of the variable scores, restandardized, so that a coefficient reads per standard deviation of the index.

How the averaging is weighted

Equal weighting of standardized components is the default and a robust one (Dawes, 1979). We keep it for judgment and for the two finance indices and depart from it for language alone. The language variables are very unevenly distributed across the nine families: tone and emotion holds 73 of the 192, 38 percent of the index, while substance and time reference holds 5. A flat average over variables would therefore be close to a tone index, and would say more about which variables

the dictionaries happen to measure often than about how founders speak. We average within family first and then across the nine families, so each family carries the same weight however many variables it holds.

Neither of the other two sets raises the problem. The judgment variables are distributed far more evenly across their nine groups; the largest holds 33 of 145 variables, 22.8 percent, against language's 38 percent, and the boundaries between those groups are a matter of judgment rather than a property of the instrument, so we impose no grouping on them. The finance indices have 36 and 11 columns and no subgroup structure beyond the hard and soft split that already defines them.

Reading a coefficient

All four indices take their signs by cross-fitting on the handshake, so a higher value means a more handshake-like pitch on that kind of information rather than more of any particular feature. We read them as composites of predictive content, not as scales of a trait, and in the text we say what an index is associated with rather than what it is. In particular, the sign of a coefficient cannot be read off the index's name. Soft finance is the case where this matters: its fitted direction runs against the ask, so a pitch scores high on it by asking for a modest sum at a modest valuation, and a positive coefficient on that index says that the smaller claim is the one more likely to reach the outcome. Take two pitches. The first asks $50,000 and projects $200,000 of revenue for next year; the second asks $1 million and projects $20 million. It is the first that scores higher on the index as the data builds it.

Direction therefore has to be read from the variables rather than from the indices, and we read it there throughout. Table 7 enters each financial figure on its own, in its own units, with no sign applied, and Appendix Table A.7 does the same for every one of the 365 variables in the four sets. Those tables are where a statement about which way something runs is established; the index coefficients carry magnitude and significance.

Because every sign is estimated, all four indices are generated regressors (**?**). The standard errors reported throughout are clustered on the episode and do not carry the uncertainty of building the index itself.

We report average marginal effects in percentage points throughout. A coefficient is the change in the probability of the outcome, averaged over the pitches in the estimation sample, for a one-standard-deviation increase in the index. The base rates are 50.5 percent for *interest*, 66.8 percent for the *handshake* and 60.7 percent for *completion*, close enough that effects are comparable across the three stages.

The four indices are weakly correlated with one another. Language and judgment correlate at 0.39, hard finance and judgment at 0.19, hard finance and language at 0.14, hard finance and soft finance at 0.13, soft finance and language at 0.08, and soft finance and judgment at -0.05. The largest by some way is between the two text indices, which is what one would expect of two

readings taken from the same words. Panel B of Table 3 is nonetheless the panel to read for what each kind of information contributes on its own.

5 Empirical Strategy

Our question is which kind of information matters at each stage of the funding process. We answer it in two complementary ways. Index regressions ask in which direction each information set is associated with success at a stage. Out-of-sample prediction asks how well each set, used in full, predicts the outcome of a stage for pitches the models have not seen. The regressions establish direction and the prediction exercise measures predictive strength.

The two on-air stages, *interest* and the *handshake*, are observed for all 642 pitches. *Completion* is documented for 313 of the 429 handshakes. This is too few to split into training and test samples, so completion is analyzed with the index regressions only.

5.1 Index regressions

Each information set is reduced to one index by the rule of Section 4.5. We estimate logit regressions of *investor interest* and the *handshake* on all 642 pitches, and of *completion* on the 313 documented handshakes, with the four indices entered jointly.

The controls for on-air outcomes are the eleven structural columns of Section 4: the season and the position within it, the number of presenters, whether a guest investor was seated, the composition of the founding team, and the five largest product categories. One set carries all three decisions, so that they are compared under one specification. Standard errors are clustered on the episode, of which there are 229. Section 6.9 reports the same estimates under the full hand-coded set.

The post-show outcomes cannot carry eleven. Between 41 and 83 ventures fall on the rarer side of each, and a logit needs about ten of those for every parameter it estimates (Peduzzi et al., 1996). All four therefore use the same three controls, two indicators for the founder group and the season number, so that the rows of the table can be read against each other.

We report average marginal effects in percentage points per standard deviation of each index: the change in the probability of the outcome, averaged over the pitches in the column, for a one-standard-deviation rise in the index, holding the other three fixed. On a base rate of 50.5 percent at *investor interest*, a coefficient of +7 moves a venture from 50.5 to 57.5. All four indices take their signs by cross-fitting on the handshake, by the rule of Section 4.5, so a coefficient carries a magnitude but not a direction that can be read from the index's name. The directions are in Table 7 for the financial figures and in Appendix Table A.7 for every variable in the four sets, where each enters on its own with no sign applied.

5.2 Out-of-sample prediction

The regressions show in which direction each set moves a stage, but not how well it predicts that stage for a pitch the models have not seen. For that we drop the indices and use all of the variables themselves. We partition the variables of Section 4 into five sets: Baseline (the structural controls), Hard finance (H_{finance}), Soft finance (S_{finance}), Language (S_{language}) and Judgment (S_{judgment}), plus an overall set that pools all four and asks whether the four kinds of information are complementary or substitutes. Each set enters whole, so the exercise asks what an investor relying only on that kind of information could predict the outcome. Every set carries the structural controls, so the Baseline set is the benchmark the others are measured against and a set's margin over it is what the content of the pitch adds to the circumstances of the episode.

Three design choices discipline the comparison. First, every set runs through the same pre-specified pipeline. Second, predictions come from a soft-vote ensemble of seven learners whose composition is fixed in advance: a logistic regression with an L2 penalty, logistic regressions with an L1 and an elastic-net penalty, a random forest, a gradient-boosting machine, and the XGBoost and LightGBM implementations. The models capture different kinds of relationships. The logistic regressions capture additive effects, the L1 and elastic-net penalties test whether a sparse rule suffices in our wide and partly redundant variable sets, and the tree ensembles capture thresholds and interactions without our naming them in advance. An information set that looks weak under one functional form can be informative under another, so tying the comparison to a single model would confound what the information can predict with what that model can exploit. The ensemble averages the seven predicted probabilities with equal weight; fixing an ensemble in advance prevents us from choosing a winner on the test data.

Third, no result rests on a single train-test split. We evaluate over 50 splits into 70 percent training and 30 percent test pitches. The first is the fixed split used to generate the judgment variables. Each of the other 49 is the most balanced of up to 2,000 random draws, balanced on founder group, industry, *investor interest* and season. Because the judgment variables were proposed on the first split's training pitches, the later splits reshuffle some of those pitches into the test sample, so for that set alone the 49 later splits are not out of sample in the way they are for the other three.

Within each split, all preprocessing is fit on the training pitches only and then applied unchanged to the test pitches: capping of non-binary variables at the 1st and 99th percentiles, median imputation with missingness dummies, log transformation of skewed variables, standardization, and conversion of categorical variables into indicators. Variables that are constant in the training sample are dropped. Variables whose absolute correlation with the outcome in the training sample is below 0.02 are also dropped; this screen uses the outcome, but only in the training sample. Finally, of every pair of variables correlated at $|r| \geq 0.8$, one is removed. Each test pitch is scored once per split.

We report macro-F1 and AUC, averaged over the 50 splits. Accuracy would mislead because

the *handshake* outcome is imbalanced: two-thirds of pitches end in a *handshake*, so a model that always predicts one would score two-thirds accuracy while learning nothing. Macro-F1 computes the F1 score separately for each class and averages the two with equal weight, so a set scores well only by predicting both handshakes and failures. AUC measures how well a set ranks pitches by predicted probability, independent of the 0.5 cutoff that turns probabilities into predictions. Two sets are compared by their difference in AUC within each split, where both were scored on the same test pitches. Because the 50 splits are resamples of the same 642 pitches rather than independent samples, we report the number of splits in which one set is ahead as the primary summary, with Wilcoxon signed-rank tests on the paired differences in the appendix.

6 Results

This section asks one question in several ways: which of the four kinds of information is associated with which decision. It begins with the three decisions themselves and how far they overlap, puts the four indices against each one in turn, then asks what inside the judgment index carries it, what the *handshake* was priced at, what became of the ventures after the broadcast, and how much of all this survives an out-of-sample exercise in which no index is built at all.

The answer in one sentence is that the three decisions rest on different information. LLM judgment is what draws *investor interest*, and it is also what buys the founder better terms. By the *handshake* its margin has narrowed and both sets of stated figures count beside it, the founder's forward claims among them: the larger the ask and the implied valuation, the less likely the agreement. At *completion* judgment says nothing and those same forward claims are all that speaks, so the deals that fall through are the ones that were asked for most ambitiously. The aggregate language index adds nothing with any of the three decisions once judgment is in the regression, though Section 6.4 shows that pooling hides differences between its nine families.

6.1 The three decisions

Of the 642 pitches, 324 (50.5 percent) draw investor interest, that is, at least two offers, and 429 (66.8 percent) end in a handshake. The two on-air decisions overlap but are not nested: 128 handshakes follow a single offer, and 23 pitches that draw competing offers leave without one. The correlation between them is 0.56. After the show, completion is documented for 313 of the 429 handshakes, and 190 of these (60.7 percent) were completed; the remaining 123 (39.3 percent) fell through. Figure 1 shows how the three decisions narrow the sample.

6.2 The four kinds of information across the funnel

Table 3 estimates the three decisions on the four indices, each standardized, so that every number is the change in the probability of that decision, in percentage points, for a one-standard-deviation rise in the index. The two panels differ only in what enters the regression. Panel A runs one index

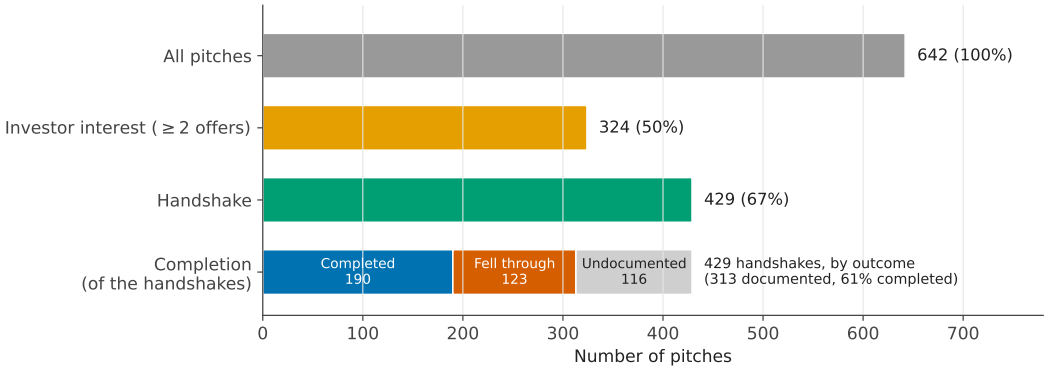


Figure 1: The three decisions

Of 642 pitches, 324 draw *investor interest* (at least two offers) and 429 end in a *handshake*; the two on-air decisions are not nested. *Completion* is documented for 313 handshakes (190 completed, 123 fell through) and undocumented for 116.

at a time, four separate regressions per column, and asks what each kind of information is worth on its own. Panel B runs all four in a single regression per column and asks what remains of each index once the others are accounted for. Panel B is the panel the claims below rest on, because the four indices are correlated with each other and a Panel A coefficient absorbs whatever an index shares with the other three. Every regression in both panels carries the eleven structural controls, with standard errors clustered on the episode. *Investor interest* and the *handshake* are estimated on all 642 pitches; *completion* on the 313 handshakes whose outcomes two independent collections established.

At the first stage, *investor interest*, the ordering is unambiguous. A standard deviation of judgment raises the probability of drawing competing offers by 16.2 percentage points ($p < 0.001$), on a base rate of 50.5 percent. The other three carry less than half of that on their own: hard finance +6.7 ($p = 0.001$), language +6.2 ($p = 0.005$), soft finance −0.4 ($p = 0.86$). Entered together, judgment keeps almost its full size, +15.4 ($p < 0.001$), hard finance holds a quarter of it, +4.0 ($p = 0.036$), and the other two go to nothing: language +0.5 ($p = 0.83$), soft finance +0.7 ($p = 0.70$). The gap between judgment and language in that model is itself significant ($\chi^2 = 15.6$, $p < 0.001$). Neither the figures the founder states nor the way they say them has a separate association with drawing offers once judgment is in the regression. The four indices are correlated, so this says that they carry no information about interest beyond what judgment already carries, not that judgment is what the earlier coefficients were measuring; a regression cannot tell those apart.

At the *handshake* judgment still leads, but by less, and the stated figures arrive beside it. Taken one at a time, judgment falls from +16.2 to +8.8 ($p < 0.001$) on a base rate of 66.8 percent, hard finance rises to +6.5 ($p < 0.001$), language holds at +4.8 ($p = 0.022$) and soft finance reaches +3.5 ($p = 0.056$). Entered together, judgment keeps its size, +8.0 ($p < 0.001$), and both finance indices stand: hard finance +4.7 ($p = 0.013$) and soft finance +3.8 ($p = 0.048$), while language drops away again (+1.3, $p = 0.55$). The distance between judgment and language narrows to the edge

Table 3: The four kinds of information across the funnel

	Investor Interest ($N = 642$)	*Handshake* ($N = 642$)	*Completion* ($N = 313$)
Panel A. One kind of information at a time			
Hard finance (H_{finance})	6.7^{**}	6.5^{***}	-0.6
	(2.0)	(1.8)	(3.0)
Soft finance (S_{finance})	-0.4	$3.5^{\dagger}$	7.9^{**}
	(2.0)	(1.8)	(2.8)
Language (S_{language})	6.2^{**}	4.8^{*}	4.3
	(2.1)	(2.1)	(3.1)
LLM judgment (S_{judgment})	16.2^{***}	8.8^{***}	1.8
	(2.0)	(1.8)	(3.2)
Panel B. All four together			
Hard finance (H_{finance})	4.0^{*}	4.7^{*}	-2.2
	(1.9)	(1.9)	(3.0)
Soft finance (S_{finance})	0.7	3.8^{*}	8.3^{*}
	(1.9)	(1.9)	(3.1)
Language (S_{language})	0.5	1.3	2.7
	(2.2)	(2.2)	(3.1)
LLM judgment (S_{judgment})	15.4^{***}	8.0^{***}	2.4
	(2.2)	(2.0)	(3.2)

Notes. Logit average marginal effects in percentage points, with standard errors in parentheses: the change in the probability of the outcome, averaged over the pitches in the estimation sample, for a one-standard-deviation increase in the index. Base rates are computed separately for each decision on its own column: 50.5 percent for *investor interest*, 66.8 for the *handshake* and 60.7 for *completion*. Every column carries the same eleven structural controls: the season and the position within it, the number of presenters, whether a guest investor was seated, two indicators for the composition of the founding team, and the five largest product categories. Standard errors are clustered on the episode. All four indices are built by the same rule: each variable is standardized and multiplied by a sign fitted on the folds a pitch is not in, so a positive coefficient means the index is higher for pitches resembling those that reached a *handshake* and the direction cannot be read from the index's name. This matters most for soft finance, whose fitted direction runs against the ask, so that a high value is a modest ask at a modest valuation. The direction of every underlying variable is in Appendix Table A.7. The language index weights its nine families equally; the other three weight their variables equally.
Interest is at least two competing offers and *Handshake* an agreement reached on air, both over all 642 pitches; *Completion* is whether the agreed investment closed, over the handshakes whose outcome two independent web collections established.
$^{\dagger}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

of significance ($\chi^2 = 3.4$, $p = 0.067$). What the forward claims say here is read from the figures beneath them, in Table 7: the ask runs -4.8 ($p = 0.012$) and the implied valuation -3.1 ($p = 0.11$), so holding the realized figures and both text indices fixed, a founder who asks for more on a higher implied valuation is less likely to leave with an agreement.

At *completion*, among the 313 *handshakes* whose *completion* two independent collections establish, what won the agreement on air says nothing about whether it was kept. LLM judgment is $+1.8$ ($p = 0.57$) on its own and $+2.4$ ($p = 0.47$) with the others; hard finance -0.6 ($p = 0.84$) and -2.2 ($p = 0.48$); language $+4.3$ ($p = 0.17$) and $+2.7$ ($p = 0.38$). The one index that speaks is soft finance, $+7.9$ ($p = 0.007$) on its own and $+8.3$ ($p = 0.010$) beside the other three, on a base rate of 60.7 percent, and *completion* is the only stage at which the two finance coefficients differ significantly from each other ($\chi^2 = 4.5$, $p = 0.033$). The forward claims are what the index is built from, and their own directions, reported one figure at a time in Table 7, say which way it runs: the amount asked carries -10.8 percentage points ($p < 0.001$), the revenue projected -12.5 ($p < 0.001$)

and the implied valuation −9.1 ($p = 0.002$). The larger the figures a founder put on the future and on the terms, the more often the agreed deal fell through.

Soft finance is also the index that grows across the funnel. Beside the other three it runs +0.7, +3.8, +8.3, and because its fitted direction runs against the ask, a rising coefficient means that a modest ask on a modest valuation matters more at each successive stage: the forward claims are silent at *investor interest*, count at the *handshake*, and count most at *completion*. Judgment runs the other way, +15.4, +8.0, +2.4, largest where the decision is cheapest and indistinguishable from nothing where it is paid for. Figure 2 traces the two paths.

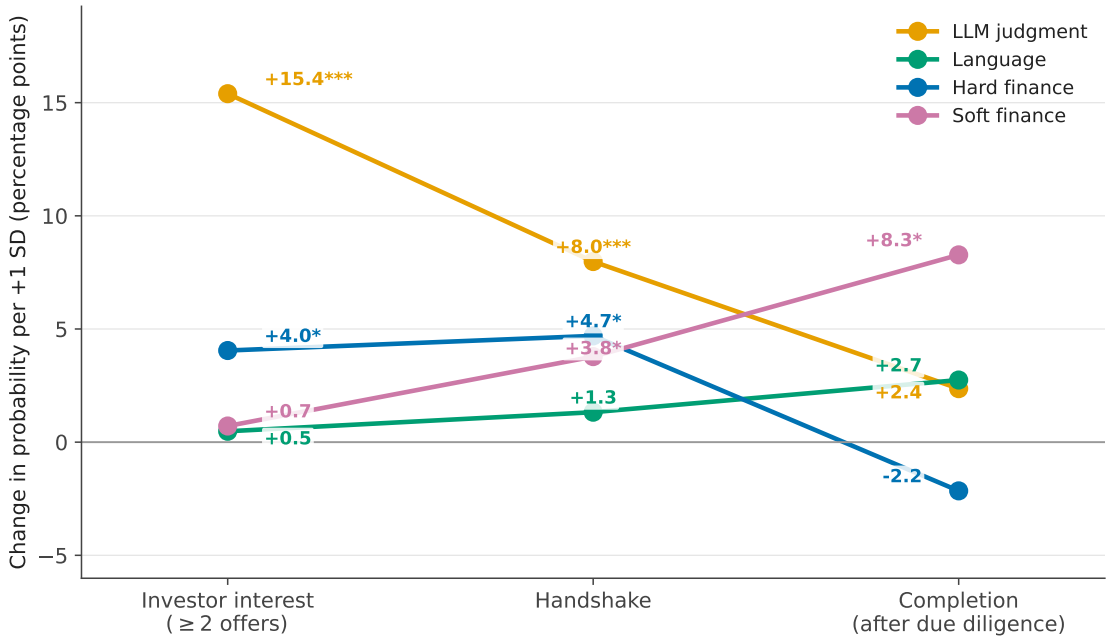


Figure 2: The four indices across the funnel

Notes. The same estimates as Table 3, Panel B: the average marginal effect of a one-standard-deviation higher index on the probability of each decision, in percentage points, from a logit with all four indices and the eleven structural controls. *Investor interest* and the *handshake*: 642 pitches; *completion*: 313 documented *handshakes*. $^{\dagger}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

A *handshake*, though, is not a single thing. The founders who reach one do not all reach the same one, and Section 6.6 asks which of these four kinds of information moves the terms they end up with.

6.3 Decomposing the LLM judgment index by variable group

The judgment index pools 145 variables, and Table 4 asks which of them carry it. The nine variable groups of Section 4 partition the set, and the table reports each group in two ways for both outcomes judgment index predicts *investor interest* and *handshake*. For each outcome, the first column enters that group's index on its own and gives the direction and size of its association, in percentage points. Entered one at a time, eight of the nine are significant. The second column

allocates the whole to these nine groups. It is the group's Shapley value, its average marginal contribution to the judgment index's gain in McFadden's pseudo R^2 over the structural controls, taken over all $2^9 = 512$ subsets of the nine groups, expressed as a percentage. These sum exactly to the index's total gain, 0.097 at *investor interest* and 0.049 at the *handshake*. There must be a gain to divide, so the decomposition covers the two on-air stages: at *completion* the judgment index itself carries +1.8 percentage points ($p = 0.57$).

Table 4: Decomposing the LLM judgment index

		Investor interest		Handshake	
Variable group	Variables	pp	Share	pp	Share
Business model and economics	16	$+15.29^{***}$	29.7%	$+9.46^{***}$	32.2%
Traction and customers	33	$+13.63^{***}$	16.4%	$+5.71^{**}$	5.1%
Capital and deal	10	$+8.85^{***}$	12.7%	$+5.11^{**}$	11.3%
Market and competition	9	$+10.05^{***}$	11.0%	$+3.67^{\dagger}$	2.3%
Operations and risk	18	$+11.86^{***}$	10.6%	$+9.74^{***}$	36.6%
Go-to-market and distribution	19	$+11.63^{***}$	8.6%	$+5.49^{**}$	4.0%
Product and defensibility	15	$+8.45^{***}$	6.8%	$+4.70^{**}$	4.1%
Team and founders	19	$+7.57^{**}$	2.9%	$+4.49^{*}$	2.2%
Mission	6	+1.32	1.3%	+1.58	2.1%
Total	145		100%		100%

Notes. pp: the logit average marginal effect in percentage points per standard deviation of that group's index, entered one at a time with the eleven structural controls and standard errors clustered on the episode, on all 642 pitches; it gives the direction and size of the association. It cannot be read as a share, because the groups overlap: entered one at a time, every group except mission is significant at both on-air stages, and their contributions would sum to far more than the whole.
Share: the group's Shapley value as a percentage of the judgment index's total gain in McFadden's pseudo R^2 over the controls, its average marginal contribution over all $2^9 = 512$ subsets of the nine groups, with the controls always in the regression. The Shapley values sum exactly to that gain, which is 0.097 at *investor interest* and 0.049 at the *handshake*. Shares carry no sign; the percentage points give the direction. Variables: the nine groups partition the 145 variables.
$^{\dagger}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

What Claude and GPT both read about the business dominates at both stages. Business model and economics takes 29.7 percent of what the index explains at *investor interest* and 32.2 percent at the *handshake*, on sixteen variables. Beyond business model and economics the two stages differ. At *investor interest* the signal is spread across what the venture has achieved and is asking for: traction and customers 16.4 percent, capital and deal 12.7, market and competition 11.0, operations and risk 10.6, with the top three groups accounting for 58.8 percent between them. At the *handshake* it concentrates on execution. Operations and risk rises from 10.6 to 36.6 percent and, with business model and economics, accounts for 68.8 percent of the total, while traction falls from 16.4 to 5.1 and go-to-market from 8.6 to 4.0. Competing offers follow an assessment of what a venture has done; the agreement follows an assessment of whether it can be run. Mission is the one group that carries nothing at either stage, +1.3 and +1.6 percentage points, neither significant: what the models read about why the venture exists is not what moves either decision.

6.4 Decomposing the language index by family

The aggregate language index adds nothing once LLM judgment is in the regression as shown in Table 3. This section asks which parts of it carry what little it has on its own, and what exactly judgment takes away.

Table 5 divides the index into its nine families. Each is entered alone as an index, and each is also given a Shapley share of the language index's gain in pseudo R^2 over the controls, computed as for judgment variable groups: the average amount that family adds across every order in which the nine could be entered, so that the nine family shares sum to the whole. The results for two outcomes barely overlap. At *investor interest* the signal is concentrated in concreteness and numeracy: ten variables counting numbers, dollars and percentages take 41.7 percent of what the index explains, and it is one of only two families significant on its own (+5.5 percentage points, $p = 0.032$), the other being substance and time reference (+4.3, $p = 0.045$). At the *handshake* concreteness collapses to 1.1 percent and the weight moves to warmth and competence (28.8), entrepreneurial orientation (23.7) and substance and time reference (23.4), 76 percent together, with warmth the only one significant (+4.1, $p = 0.047$).

Table 5: Decomposing the language index by family

		Investor interest		Handshake	
Family	Vars	pp	Share	pp	Share
Concreteness and numeracy	10	$+5.48^{*}$	41.7%	+1.30	1.1%
Substance and time reference	5	$+4.31^{*}$	20.4%	$+3.61^{\dagger}$	23.4%
Certainty and hedging	21	−2.57	12.6%	+0.26	0.1%
Entrepreneurial orientation	19	+2.34	9.4%	$+3.34^{\dagger}$	23.7%
Warmth and competence	27	+2.50	7.1%	$+4.09^{*}$	28.8%
Narrative and reasoning	10	+2.06	3.4%	+1.69	2.1%
Self-presentation and address	10	+1.55	2.6%	+2.89	12.2%
Fluency and complexity	17	+1.34	1.4%	−1.27	7.5%
Tone and emotion	73	+1.08	1.4%	−0.23	1.0%
Total	192		100%		100%

Notes. pp: the logit average marginal effect in percentage points per standard deviation of that family's index, entered one at a time with the eleven structural controls and standard errors clustered on the episode, on all 642 pitches. Share: the family's Shapley value as a percentage of the language index's total gain in McFadden's pseudo R^2 over the controls, its average marginal contribution over all $2^9 = 512$ subsets of the nine families, with the controls always in the regression. The shares sum exactly to that gain, which is 0.017 at *investor interest* and 0.019 at the *handshake*, against 0.097 and 0.049 for LLM judgment, so these are shares of a much smaller total. Shares carry no sign; the percentage points give the direction. The nine families partition the 192 language variables and are the units the index is weighted by (Section 4).
$^{\dagger}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

Tone and emotion account for 73 of the 192 language variables, 38 percent of the index, and contribute 1.4 percent of what language explains at *investor interest* and 1.0 percent at the *handshake*, carrying +1.1 and −0.2 percentage points on their own. The index weights the nine families equally rather than the variables, and the reason is the size imbalance itself and not this result: a flat average over 192 variables would be three-eighths a tone index whatever tone turned out to predict. We note the inertness because it is informative about the instrument, not because it

justified the weighting. It is also the family a reader would expect to matter, and it does not: whether the pitch sounds positive or negative is not what moves either decision.

Table 6 turns to what judgment absorbs. Each row is two logits on all 642 pitches with the eleven structural controls and standard errors clustered on the episode. The first puts one family index on the right-hand side, the same regression as Table 5 reports. The second adds the judgment index of Table 3, entire and unchanged, as one further regressor. The r column is the plain correlation between the family index and the judgment index. The two indices share no variables: language is 192 dictionary and parser counts, judgment is 145 variables the models scored, and what they have in common is the transcript both are computed from.

Judgment overlaps with one family far more than the rest: its correlation with concreteness and numeracy is 0.47, against 0.22 for substance, 0.11 for warmth, 0.02 for entrepreneurial orientation and 0.01 for fluency. Those correlations predict almost exactly which families survive. Entering the judgment index alongside each family in turn, concreteness falls from +5.5 to −1.7 percentage points at *investor interest* and from +1.3 to −3.1 at the *handshake*, and narrative and reasoning likewise crosses zero; substance keeps about half its effect and warmth rather more, though neither keeps its significance. The families judgment is uncorrelated with lose nothing at all: entrepreneurial orientation goes from +2.3 to +2.7 at *investor interest* and from +3.3 to +3.4 at the *handshake*, and fluency from +1.3 to +1.2.

Table 6: What LLM judgment absorbs from each language family

		Investor interest		*Handshake*	
Family	r with judgment	Alone	With judgment	Alone	With judgment
Concreteness and numeracy	+0.47	+5.48*	−1.74	+1.30	−3.07
Substance and time reference	+0.22	+4.31*	+1.79	+3.61†	+2.18
Narrative and reasoning	+0.20	+2.06	−0.86	+1.69	+0.16
Tone and emotion	+0.12	+1.08	+0.18	−0.23	−0.77
Warmth and competence	+0.11	+2.50	+0.85	+4.09*	+3.24
Self-presentation and address	+0.10	+1.55	+1.02	+2.89	+2.59
Certainty and hedging	+0.04	−2.57	−2.69	+0.26	+0.21
Entrepreneurial orientation	+0.02	+2.34	+2.66	+3.34†	+3.43†
Fluency and complexity	+0.01	+1.34	+1.15	−1.27	−1.35

Notes. r: correlation between that family's index and the LLM judgment index. The other four columns are the family's average marginal effect in percentage points per standard deviation, from a logit with the eleven structural controls and standard errors clustered on the episode, first with the family index alone and then with the LLM judgment index added. Families are ordered by their overlap with judgment.
$^{\dagger}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

The mechanism is straightforward once stated. Concreteness and numeracy measure a founder putting figures into the pitch, which is the same material the models read when they assess the venture, so judgment subsumes it and then some. Warmth, entrepreneurial orientation and fluency sit further from what judgment captures, and they are untouched. Hence, what judgment displaces is the numerate language, not the warm language.

What survives judgment does not rescue language as an information set. Everything the nine families explain together is a sixth of what judgment explains at *investor interest* and two-fifths

at the *handshake*, so even the family that survives best is working with very little: warmth and competence take 28.8 percent of language's total at the *handshake*, which is about a ninth of what judgment explains at the same stage. How much of that smallness belongs to language and how much to the instrument we cannot say. A dictionary counts words without knowing what they are doing in the sentence, so "warmth" here is a rate of warm-sounding words rather than a reading of a warm founder, and what we measure is the part of language that survives being counted.

6.5 Inside the finance indices

The two finance indices are small enough to open up variable by variable, and doing so shows which figure is doing the work at each stage. An index coefficient is an average over its figures, and the average can hide a handover. Table 7 reports each variable twice. Every variable enters a regression of its own, standardized, with the eleven structural controls, and the first column is its effect in percentage points. The second is its Shapley value: its average contribution to the index's gain in McFadden's pseudo R^2 over the same controls. The shares sum to the whole. Soft finance has seven figures and the decomposition is exact over all 128 subsets; hard finance has twenty-one, so its shares are estimated over 400 random orderings with a Monte Carlo error of at most 0.3 points.

Each index has a different figure in charge at each stage, and the handovers are sharp. In hard finance, *investor interest* runs on current revenue: it takes 48.1 percent of what the index explains and is the strongest single figure in it (+11.5 percentage points, $p < 0.001$), with profitability second (22.6 percent, +7.6, $p < 0.001$). At the *handshake* the two trade places. Profitability takes 44.5 percent at +8.0 ($p < 0.001$), while current revenue falls to 6.6 percent and +3.6. Having revenue is what draws the offers; being profitable is what turns them into an agreement. At *completion* neither speaks. The hard finance shares scatter instead across figures few founders state: personal liability takes 21.3 percent on 135 stated values, with a marginal effect of +7.1 ($p = 0.093$), and past revenue growth 14.3 percent on 180, not significant. The index itself carries nothing at that stage, and this is the reason.

Soft finance hands over twice. At *investor interest* the weight sits on the equity offered, 39.7 percent, and the direction is negative: a founder who offers a larger stake draws *less* interest (−7.3, $p < 0.001$). At the *handshake* the size of the ask takes over at 43.2 percent, also negative (−4.8, $p = 0.012$). At *completion* it is expected revenue, 44.8 percent at −12.5 ($p < 0.001$), with the ask at 22.6 percent (−10.8, $p < 0.001$) and the implied valuation at 14.3 (−9.1, $p = 0.002$). Those three carry 81.7 percent of the index, and the single number a founder puts on next year's revenue is the largest identifiable quantity behind whether an agreed deal closes.

Two of those figures do not merely change weight across the funnel; they change sign. Expected revenue is +5.5 percentage points at *investor interest* ($p = 0.011$) and −12.5 at *completion* ($p < 0.001$). The implied valuation is +5.2 at *investor interest* ($p = 0.014$) and −9.1 at *completion* ($p = 0.002$). The same figure that goes with a venture drawing competing offers goes with the resulting agreement being less likely to be kept, and in both cases the reversal is significant at each

Table 7: Inside the finance indices: direction and Shapley share, variable by variable

		Interest		Handshake		Completion	
	Stated	pp	Share	pp	Share	pp	Share
Panel A. Soft finance							
Equity offered	642	-7.3^{***}	39.7%	−0.3	7.7%	+2.2	3.8%
Expected revenue	279	$+5.5^{*}$	30.9%	−1.2	1.5%	-12.5^{***}	44.8%
Implied valuation	642	$+5.2^{*}$	20.2%	−3.1	20.3%	-9.1^{**}	14.3%
The ask	642	+2.5	7.9%	-4.8^{*}	43.2%	-10.8^{***}	22.6%
Payback period	7	−0.8	0.9%	−1.7	4.3%	+0.4	0.9%
Market size, customers	30	+0.7	0.3%	+2.9	21.5%	+1.9	1.4%
Market size, revenue	45	+0.3	0.1%	+0.5	1.3%	$-5.4^{\dagger}$	12.3%
total pseudo R^2 gain			0.019		0.015		0.082
Panel B. Hard finance							
Current revenue	595	$+11.5^{***}$	48.1%	$+3.6^{*}$	6.6%	−2.8	1.8%
Profitable	295	$+7.6^{***}$	22.6%	$+8.0^{***}$	44.5%	+3.2	3.3%
Units sold	140	$+4.3^{\dagger}$	6.2%	$+3.2^{\dagger}$	4.5%	+0.3	0.2%
Bootstrapped	571	+3.2	5.9%	+0.7	1.5%	+3.1	2.1%
Prior investment	602	−1.6	3.4%	−0.5	0.8%	−3.5	2.7%
Gross margin	493	+2.8	2.8%	$+3.5^{\dagger}$	5.5%	+0.7	1.0%
Debt outstanding	66	−3.1	2.5%	−2.8	2.9%	+4.8	12.8%
Development expense	47	+1.7	1.6%	+0.0	0.4%	+1.7	1.9%
Funding raised	147	−2.2	1.4%	−1.7	1.0%	+0.1	0.2%
Any marketing spend	284	−0.8	1.4%	−1.5	0.6%	−2.1	0.7%
Fixed costs	22	+1.5	1.2%	+2.6	5.5%	−2.3	0.8%
Unit price	580	+0.9	0.8%	−3.3	4.3%	$-4.9^{\dagger}$	9.9%
Outsider equity share	109	−1.5	0.5%	+0.2	0.4%	−3.9	7.2%
Founder capital	145	−1.4	0.5%	−2.8	2.8%	+4.9	7.3%
Marketing spend	71	+0.4	0.4%	−4.1	6.8%	−0.3	0.2%
Carries debt	366	−1.0	0.3%	-3.5^{*}	8.3%	+0.4	5.7%
Unit cost	477	−0.2	0.3%	−2.1	1.3%	−1.8	3.6%
Gross margin, stated	264	+0.3	0.1%	+0.8	0.3%	+1.1	1.0%
Acquisition cost	67	−0.4	0.1%	−1.8	0.6%	+2.2	2.0%
Personal liability	135	−0.3	0.0%	−0.6	0.9%	$+7.1^{\dagger}$	21.3%
Past revenue growth	180	−0.2	0.0%	−1.1	0.4%	+5.8	14.3%
total pseudo R^2 gain			0.054		0.049		0.076

Notes. Stated: the number of pitches that state the figure; the rest receive the imputed value, so a share resting on few stated values should be read with care. pp: the logit average marginal effect in percentage points per standard deviation of that variable entered alone, with the eleven structural controls and standard errors clustered on the episode. Each variable is in its own units here rather than inside an index, so the sign is the finding: a positive effect means a larger stated figure goes with the outcome.
Share: the variable's Shapley share of the index's gain in McFadden's pseudo R^2 over the same controls, its average contribution over orderings in which it is added, so that variance two variables share is split between them rather than awarded to whichever enters first. Shares within a panel sum to 100 percent of the gain printed beneath, and carry no sign. Panel A averages over all $2^7 = 128$ subsets of its seven variables exactly; Panel B samples 400 random orderings of its twenty-one, with a Monte Carlo standard error of at most 0.3 points. The not-stated indicators are part of each index but not of this table, which decomposes the indices into the figures themselves.
$^{\dagger}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

end. A large forecast and a high valuation are associated with greater interest in the room and with lower completion afterwards. This is also what the soft finance index is reporting at the two ends of the funnel, and the decomposition says which figures it is reporting it from.

6.6 The terms of the agreement

A *handshake* settles that an investment will happen, not on what terms. Among the 429 *handshakes*, the two can be separated, because every pitch opens with an ask, a sum of money for an equity share of the company, and every agreement closes on a sum of money and an equity share that can be set against it.

The comparison is one-sided. In 354 of the 429 agreements the founder gets exactly the money asked for; the median difference is zero. The equity moves instead: 374 founders hand over more of the company than they offered and 14 hand over less, and the median founder who offers ten percent walks out having sold twenty. Founders keep the headline number and pay for it in ownership.

Table 8 puts the four kinds of information against both quantities, adding one at a time. Panel A is the share of the company agreed minus the share offered, in percentage points of the company. Panel B is the money agreed minus the money asked, in dollars. The median pitch asks for $250,000 in exchange for ten percent of the company, a valuation of $2.5 million, which is the scale against which both panels should be read. Both are winsorized at the first and ninety-ninth percentile, as the stated financial figures are elsewhere in the paper, and the level of the ask is controlled for in every column. Each coefficient is the move in the outcome for one standard deviation along that kind of information. Soft finance here is the four forward claims only: expected revenue, the market size in customers and in revenue, and the payback period. Its other three figures are the ask, the equity offered and the valuation they imply, and those are the quantities the outcome is built from.

LLM judgment is what keeps a founder's equity. Because the outcome in Panel A counts what the founder hands over beyond the share offered, a negative coefficient means the founder keeps more of the company. LLM judgment is what keeps it. A founder one standard deviation higher on LLM judgment hands over 2.7 points less of the company, which would take the median case from selling twenty percent of the business to selling about seventeen. The estimate barely moves as the other kinds of information are added: 2.4 points on its own, 2.7 next to language, 2.7 with the realized figures, 2.7 with the forward claims as well. Neither finance set does anything to the terms, and the two together raise the explained variation from 0.160 to 0.163.

Language runs the other way, and only shows up once LLM judgment is in the regression. On its own it is indistinguishable from zero (+0.31); next to LLM judgment, one standard deviation of language means handing over about one point more. The two are correlated at 0.39, so the reading could be that a polished pitch not backed by a working business gets paid for in ownership. It is much the weaker of the two findings: it is significant only at ten percent, so it should be read as suggestive rather than established.

Panel B turns to the money. Nothing moves the gap between the amount agreed and the amount

Table 8: What moves the terms

	(1)	(2)	(3)	(4)	(5)
Panel A. Equity handed over minus equity offered, in percentage points					
LLM judgment		-2.389^{***}	-2.690^{***}	-2.703^{***}	-2.670^{**}
		(0.704)	(0.744)	(0.808)	(0.821)
Language	+0.311		$+0.952^{\dagger}$	$+0.949^{\dagger}$	$+1.006^{\dagger}$
	(0.525)		(0.545)	(0.546)	(0.529)
Hard finance				+0.054	+0.056
				(0.482)	(0.484)
Soft finance					−0.487
					(0.541)
R^2	0.122	0.154	0.160	0.160	0.163
N	428	428	428	428	428
Panel B. Amount agreed minus amount asked, in dollars					
LLM judgment		−7,133	−6,887	−6,994	−6,355
		(8,471)	(9,359)	(9,741)	(9,962)
Language	−2,440		−785	−813	+293
	(6,798)		(7,603)	(7,594)	(7,441)
Hard finance				+419	+453
				(5,596)	(5,598)
Soft finance					−9,379
					(7,018)
R^2	0.089	0.091	0.091	0.091	0.099
N	429	429	429	429	429

Notes. Ordinary least squares. Each index is standardized, so a coefficient is the move in the outcome for a one standard deviation move along that kind of information. Panel A is in percentage points of the company, Panel B in dollars. Both outcomes are winsorized at the first and ninety-ninth percentile of the estimation sample, the same rule applied to the stated financial figures. Every column carries the eleven structural controls and the level of the ask, the equity offered and the valuation they imply. Standard errors in parentheses are clustered on the episode (208 clusters). Soft finance is the four forward claims: expected revenue, market size in customers, market size in revenue, and the payback period. Its remaining three figures are the ask, the equity offered and the valuation they imply, which are the quantities the outcome is built from.
$^{\dagger}\,p < 0.10$, $^{*}\,p < 0.05$, $^{**}\,p < 0.01$, $^{***}\,p < 0.001$.

asked. Not language, not LLM judgment, not the realized figures, not the forward claims: every coefficient in Panel B is smaller than its own standard error. With five pitches in six settling at exactly the ask there is almost nothing there to move, and the right reading of the panel is that founders do not negotiate on the headline number rather than that these four kinds of information fail to explain something.

Which part of LLM judgment does the work is not the part that moves the decisions. Table A.3 in Appendix A.3 takes the nine variable groups of Section 6.3 one at a time and puts the terms beside the two decisions. Eight of the nine groups move *investor interest* and the *handshake*. Four move the terms, and they are the four about whether the business is working: traction and customers, the business model, going to market, operations and risk. The other five move the decisions and leave the terms alone: bundled, they carry +13.1 and +7.1 percentage points on the two decisions and −0.67 points of equity on the terms, where the four working groups carry −2.54.

So the three decisions and the terms are not the same question. Getting an investor interested rewards everything a model can read off a pitch. Being priced well rewards only the part that says the business is already working.

6.7 Past the deal

The funnel described above ends at completion. Whether a venture was still operating three years later, whether it raised capital again, whether it was acquired, are not outcomes the *Shark Tank* investors chose on air. This section asks whether anything the four indices read off the pitch is still visible in the record of the venture years later, and in particular whether the index that moved the investors most is the one that tracks what the business became. Outcomes are taken from the open-web collection described in Section 2.

Table 9 estimates four post-show outcomes on the four indices, each standardized, so that every number is again the change in the probability of the outcome in percentage points, for a one-standard-deviation rise in the index. The three panels differ only in what enters the regression. Panel A runs one index at a time. Panel B runs all four in a single regression. Panel C runs all four together with the on-air *handshake* held fixed, which removes the ventures' having received an investment as an explanation of what followed. The post-show outcomes cover the sample unevenly, and each regression below states the sample it is estimated on: operating status is established for 593 ventures, survival to year three for 498 (Seasons 1–14 only), a new investor round for the 371 ventures where the collection establishes it, and acquisition for 587.

LLM judgment reaches past the stage, and it is the only index that does so at more than one outcome. A standard deviation raises the probability that a venture is still operating at the collection date by 5.0 percentage points ($p < 0.001$) on a base rate of 86.0 percent, that it is alive three years after airing by 4.9 ($p < 0.001$) on 90.8, and that it raised a new investor round by 4.5 ($p = 0.014$) on 15.4. All three hold with all four indices in the regression (Panel B), +5.3, +5.3 and +5.6, and with the *handshake* added to them (Panel C), +5.2, +5.1 and +5.7. Whatever the models read in the pitch, it is not a reading of what the investors were about to do: the *handshake* itself carries +1.0 ($p = 0.74$) on operating status and +2.8 ($p = 0.27$) on survival to year three. A high value is a favorable reading of the business, not a large number in it.

Hard finance reaches past the stage too, in two outcomes, and in opposite directions. It raises the probability of surviving to year three, +2.7 ($p = 0.019$), holding at +2.2 with all four indices in the regression ($p = 0.072$), while operating status is flat (+1.7, $p = 0.20$). And it lowers the probability of raising a new round, by much the largest coefficient in the table: −8.7 on its own ($p < 0.001$), −8.4 with the other three and −8.2 with the *handshake* added. The two signs are consistent rather than contradictory. A venture whose realized figures read well is a business that works, so it is more likely to be operating later; and it is a business with less need of outside money, so it is less likely to have raised again. What looks like a negative result for strong finances is a statement about demand for capital, not about survival.

Table 9: The four kinds of information past the deal

	Operating at collection ($N = 593$)	Alive at year three ($N = 498$)	New funding round ($N = 371$)	Acquisition ($N = 587$)
Panel A. One kind of information at a time				
Hard finance (H_{finance})	$+1.7$	$+2.7^{*}$	-8.7^{***}	-0.0
	(1.3)	(1.2)	(2.0)	(1.0)
Soft finance (S_{finance})	-0.8	-1.0	-5.9^{*}	$+0.4$
	(1.1)	(1.0)	(2.6)	(0.7)
Language (S_{language})	$+0.2$	$+0.2$	$+0.1$	$+0.5$
	(1.3)	(1.4)	(1.9)	(1.1)
LLM judgment (S_{judgment})	$+5.0^{***}$	$+4.9^{***}$	$+4.5^{*}$	$+2.5^{\dagger}$
	(1.3)	(1.2)	(1.9)	(1.3)
Panel B. All four together				
Hard finance (H_{finance})	$+1.2$	$+2.2^{\dagger}$	-8.4^{***}	-0.6
	(1.3)	(1.2)	(1.9)	(1.0)
Soft finance (S_{finance})	-0.5	-0.9	$-4.6^{\dagger}$	$+0.6$
	(1.2)	(1.1)	(2.5)	(0.7)
Language (S_{language})	-1.6	-1.8	-0.5	-0.4
	(1.3)	(1.4)	(1.9)	(1.1)
LLM judgment (S_{judgment})	$+5.3^{***}$	$+5.3^{***}$	$+5.6^{**}$	$+2.7^{\dagger}$
	(1.4)	(1.3)	(2.1)	(1.4)
Panel C. All four together, controlled for the handshake				
Hard finance (H_{finance})	$+1.1$	$+2.0$	-8.2^{***}	-0.7
	(1.3)	(1.2)	(1.9)	(1.0)
Soft finance (S_{finance})	-0.5	-1.0	$-4.6^{\dagger}$	$+0.6$
	(1.2)	(1.1)	(2.6)	(0.7)
Language (S_{language})	-1.7	-1.9	-0.5	-0.4
	(1.3)	(1.4)	(1.9)	(1.1)
LLM judgment (S_{judgment})	$+5.2^{***}$	$+5.1^{***}$	$+5.7^{**}$	$+2.5^{\dagger}$
	(1.4)	(1.3)	(2.1)	(1.4)
The *handshake* itself	$+1.0$	$+2.8$	-2.3	$+2.5$
	(3.0)	(2.7)	(4.2)	(2.5)

Notes. Logit average marginal effects in percentage points per standard deviation of each index, with standard errors beneath, clustered on the episode. Every outcome is taken from the open-web collection: a value is retained only where two independent collection passes agree, which is why the four samples differ and why no column is restricted to the intersection. Operating at collection counts a venture active or acquired against closed; a new funding round is a priced round from an outside investor. The ones in each column are 510, 452, 57 and 41, so the base rates are 86.0, 90.8, 15.4 and 7.0 percent, and a coefficient should be read against the base rate of its own column.
Every column carries the same three controls: two indicators for the founder group and the season number, standardized. Panel C is Panel B with the *handshake* added as a dummy, whose own effect is the discrete difference between a venture that shook hands and one that did not.
$^{\dagger}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

Language predicts nothing anywhere. Its four coefficients are +0.2, +0.2, +0.1 and +0.5 percentage points, each smaller than its own standard error, and all four stay insignificant when judgment is in the regression. The index that drew competing offers on its own at the first stage, and then dropped away once judgment was accounted for, leaves no trace at all in what the venture became.

Soft finance speaks at one outcome only, the new funding round, where it carries −5.9 percentage points ($p = 0.027$) on a base rate of 15.4 percent and weakens to −4.6 in the other two panels ($p = 0.072$ and $p = 0.078$). On survival and acquisition, it is null, −0.8, −1.0 and +0.4. Its fitted

direction runs against the ask, so a negative coefficient here says that a founder who asks for more, on a higher implied valuation and a larger forecast, is more likely to raise institutional money afterwards. The same claims that cost them the *handshake* and the completion are the ones that go with later funding, which is the sharpest instance in the paper of a signal being read differently in the room and in the market.

The sharpest version of this question is what happens among the ventures the investors did not back. If judgment were a reading of what the investors were about to do, it would have nothing to say about a venture that received no investment. Table 10 splits the outcomes the data support by whether the pitch ended in a handshake. Among the ventures that did not reach one, a standard deviation of judgment raises the probability that the venture is still operating by +6.9 percentage points ($p = 0.008$, on 195 ventures) and that it survived to year three by +9.0 ($p < 0.001$), against +4.3 ($p = 0.008$) and +3.0 ($p = 0.044$) among the ventures that reached one. The point estimates are larger among the pitches without a handshake, though the standard errors are nearly twice as wide, because there are half as many of them. Entered on all 642 pitches with an interaction between judgment and the handshake, the difference between the two groups is indistinguishable from zero at both outcomes ($p = 0.87$ and $p = 0.50$). Judgment reads the venture, and it reads it whether or not an investor agreed to put money into it.

The pattern is not uniform. Raising a new round is the one outcome where the two groups look different: judgment carries +6.5 among the ventures that reached a *handshake* ($p = 0.003$) and +3.9 among the rest ($p = 0.33$), and the interaction is again insignificant ($p = 0.91$), so the gap is not estimated precisely enough to interpret. The realized figures carry that outcome in both groups, −10.2 ($p = 0.008$) among among the pitches without a *handshake* and −7.8 ($p < 0.001$) among those with one. Language carries nothing in either group at any outcome. Acquisition is left out of the table: eleven of the 194 pitches without a *handshake* were acquired, too few events to estimate.

One source is deliberately absent from the table. Crunchbase data was also collected alongside the open-web record and covers 388 of the 642 ventures, and appearing in it is not independent of the pitch. It is widely used in economic and managerial research (Dalle et al., 2017), and its coverage of private companies is known to favor larger transactions: benchmarking eight commercial databases against 339 verified financing rounds, Retterath and Braun (2020) find that larger rounds are more likely to be reported than smaller ones, and that round sizes and valuations are themselves more likely to be recorded for the larger rounds. Our sample shows the same selection. Crunchbase presence of the ventures in our data is strongly predicted by what the venture said about its finances: a standard deviation of the implied valuation raises it by 13.8 percentage points, the ask by 11.2 and prior investment by 10.8, while a larger equity stake offered lowers it by 12.3, being bootstrapped by 10.8 and being profitable by 8.6, all at $p < 0.001$. Together the structural observables predict presence with an out-of-sample AUC of 0.696, and LLM judgment adds 6.4 percentage points ($p = 0.012$) on top of them. The database covers the ventures that look like venture-capital deals and misses the self-funded, profitable, small-ask business, which is what such

Table 10: Past the deal, split by whether the pitch ended in a handshake

	Operating at collection		Alive at year three		New funding round	
	No handshake	Handshake	No handshake	Handshake	No handshake	Handshake
Hard finance (H_{finance})	+2.0	+1.3	+2.5	+2.0	-10.2^{**}	-7.8^{***}
	(2.7)	(1.5)	(2.7)	(1.4)	(3.4)	(2.2)
Soft finance (S_{finance})	−5.3	+0.6	−2.1	−0.6	−0.8	−5.2
	(3.4)	(1.3)	(3.4)	(0.9)	(4.1)	(3.1)
Language (S_{language})	−2.2	−1.4	−3.1	−1.2	+4.7	−3.3
	(2.4)	(1.7)	(2.7)	(1.6)	(3.5)	(2.2)
LLM judgment (S_{judgment})	$+6.9^{*}$	$+4.3^{**}$	$+9.0^{***}$	$+3.0^{*}$	+3.9	$+6.5^{**}$
	(2.6)	(1.6)	(2.6)	(1.5)	(4.0)	(2.4)
Ventures	195	398	168	330	114	257
Of which the outcome occurred	160	350	145	307	21	36
Is judgment different between the two groups? Interaction on all 642 pitches						
LLM judgment × handshake	−0.044 ($p = 0.872$)		−0.256 ($p = 0.499$)		+0.036 ($p = 0.910$)	

Notes. Logit average marginal effects in percentage points per standard deviation of each index, with standard errors beneath, all four entered together, estimated separately on the pitches that did and did not end in a handshake. Every column carries the same controls as Table 9: two indicators for the founder group and the season number, standardized. Standard errors are clustered on the episode. The interaction row is one regression on all 642 pitches with the handshake and its interaction with judgment added, and reports the interaction coefficient, so a value indistinguishable from zero means judgment is worth the same in both groups. Outcomes are taken from the open-web collection and retained only where the two independent passes agreed, which is why the number of ventures differs across columns. Acquisition is omitted: eleven of the 194 pitches without a *handshake* with a recorded outcome were acquired, too few events to estimate a logit.
$^{\dagger}\, p < 0.10$, $^{*}\, p < 0.05$, $^{**}\, p < 0.01$, $^{***}\, p < 0.001$.

a database is for. The same test run on the four open-web outcomes returns AUCs between 0.49 and 0.57 with all four indices insignificant throughout, so the open record does not select on what the indices measure. Appendix A.4 reports the selection test in full, and the two outcomes as the database records them.

6.8 Out of sample prediction

Everything so far runs on the four indices, each standardized, signed by cross-fitting and collapsed to one column. This section asks the same question of the variables themselves. Each information set enters as its own columns, with no signing, no collapsing and no index, through the pipeline of Section 5: 50 train/test splits, every step of the preparation fitted inside each split, a seven-learner soft vote read on the test pitches. The question it answers is how much of a decision any of this calls in advance, and whether the ordering of Table 3 depends on how the indices were built.

Completion is not part of the exercise. It is documented for 313 handshakes, so a test sample would hold about 90 pitches, too few for a stable estimate over 50 splits; the regressions above answer it on all 313. Every column set includes the structural controls, so each row is what that information set adds to the controls. The controls here are not the eleven of Table 3 but the full structural set of 67 encoded columns, which a penalized learner can carry where a logit on 313 completions cannot; the Baseline row is therefore a harder benchmark than Table 3's control set,

not the same one. Two sets are compared within a split, where both were scored on the same test pitches, so the comparison is unaffected by a split being easy or hard; Panel B of Table 11 reports those within-split differences directly.

Table 11: Out-of-sample prediction by information set

		Investor interest		Handshake	
Information set	Columns	AUC (SD)	Δ, wins	AUC (SD)	Δ, wins
Panel A. Each information set against the structural controls					
Baseline (structural controls only)	67	0.489 (0.039)	–	0.451 (0.039)	–
Hard finance (H_{finance})	103	0.565 (0.034)	+0.076 48/50	0.575 (0.029)	+0.125 50/50
Soft finance (S_{finance})	78	0.536 (0.036)	+0.047 48/50	0.471 (0.036)	+0.020 38/50
Language (S_{language})	259	0.547 (0.028)	+0.058 48/50	0.500 (0.035)	+0.050 44/50
LLM judgment (S_{judgment})	228	0.652 (0.035)	+0.163 50/50	0.571 (0.032)	+0.120 50/50
All four together	467	0.630 (0.034)	+0.141 50/50	0.587 (0.035)	+0.136 50/50
Panel B. Head to head, within-split difference and wins					
LLM judgment − hard finance			+0.087 50/50		−0.005 21/50
LLM judgment − language			+0.105 50/50		+0.070 46/50
Hard finance − soft finance			+0.029 39/50		+0.105 49/50
LLM judgment − all four together			+0.022 42/50		−0.016 14/50

Notes. Test-sample area under the ROC curve for a seven-learner soft vote, averaged over 50 train/test splits, with the standard deviation across splits in parentheses. Every information set carries the structural controls, so Panel A is what the set adds to the controls alone. Δ is the mean difference within a split, where both sides were scored on the same test pitches, and *wins* is the number of the 50 splits in which the first is ahead. Because the splits are resamples of the same 642 pitches rather than independent samples, the win count is more informative than a p-value; Wilcoxon signed-rank tests on the 50 paired differences are in Appendix A.5, and the only contrasts they do not separate are judgment against hard finance at the *handshake* ($p = 0.30$) and soft finance against language at *investor interest* ($p = 0.23$). Imputation, winsorization, encoding, scaling and column selection are fitted on each split's training pitches alone, which is why the column count is a mean. Test samples hold 193 pitches.

At *investor interest* the ordering is the one the regressions give. Judgment reaches an AUC of 0.652 against a baseline of 0.489, a gain of 0.163 that holds in all 50 splits. It beats hard finance by 0.087 within splits, in 50 of 50, language by 0.105 and soft finance by 0.116, each in 50 of 50. Hard finance reaches 0.565 (+0.076), language 0.547 (+0.058) from its 192 columns and soft finance 0.536 (+0.047) from its eleven, each in 48 of 50 splits.

At the *handshake* hard finance catches up, again as in the regressions: 0.575 against judgment's 0.571, a difference of 0.005 that goes either way across splits ($p = 0.30$). Language falls to 0.500, where its margin over the baseline is +0.050 and holds in 44 of 50 splits, and soft finance to 0.471, where its margin is +0.020 and holds in only 38. The two finance sets separate by 0.105 in favour of hard finance, in 49 of 50 splits. None of this uses the orientation or the pooling, so the pattern of Table 3 is not a product of either.

Combining the four does not beat the best single set at interest: 0.630 against judgment's 0.652, behind it in 42 of 50 splits. At the *handshake* it does, 0.587 against 0.571, ahead in 36 of 50 ($p = 0.010$), which is the same thing Panel B of Table 3 shows at that stage, where judgment and hard finance both survive with the other in the regression.

Figure 3 shows the spread behind these averages. The standard deviation across splits runs from 0.028 to 0.039 for every information set, a fifth to a quarter of judgment's gain at *investor*

interest, so no single split settles anything and every comparison above is made within splits. What the boxes show is the reversal itself: at *interest* judgment's distribution sits clear of the baseline and of the other three sets, while at the *handshake* it overlaps hard finance almost completely, which is the same 21 of 50 splits that Panel B records.

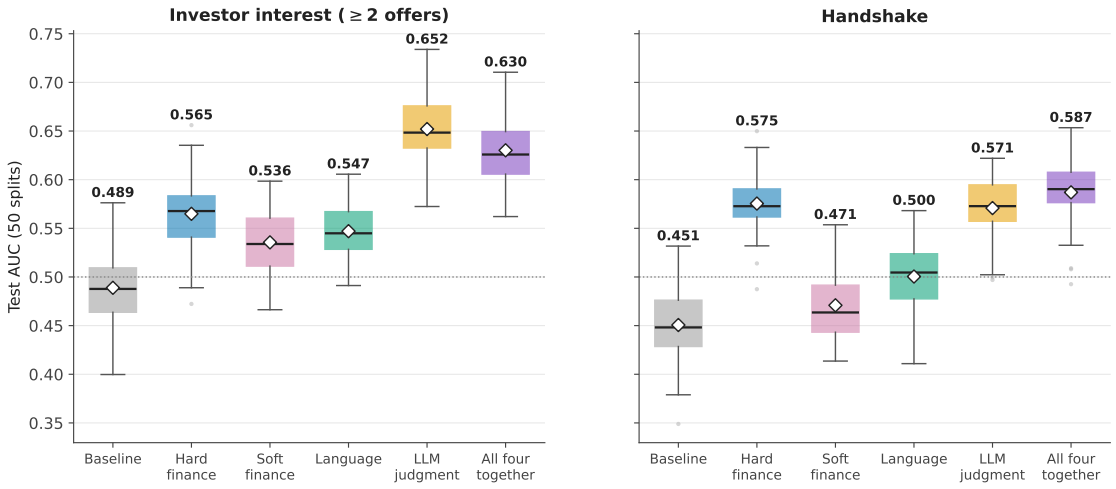


Figure 3: Spread of the out-of-sample AUC across splits

Notes. Each box is the distribution of the test AUC of the seven-learner ensemble over 50 train/test splits; the diamond and the number mark the mean; the dotted line is 0.5. Each information set enters with the structural controls. The companion figure in Appendix A.5 Figure A.1 shows the same for macro-F1.

6.9 Robustness

The paper rests on four pillars: the dataset, the LLM variables built from it, the measurement and index construction, and the empirical strategy. What the dataset can and cannot support is in Section 2.5. This section takes the other three in turn. Appendix B holds the detail.

LLM variables

The entirety of judgment variables were written and scored by language models, therefore there are four important robustness checks to conduct apart from prompts and codebooks provided in the Appendix C.2.4: i) do the scoring of existing variables reproduce; ii) do the generated list of variables reproduce; iii) whether the result depends on one provider and finally iv) whether the models are recognizing ventures they have seen rather than reading the pitch.

Scoring reproduces. The first question is whether the same scores come back when the same model reads the same transcripts again. Given the codebook, rescoring the same transcripts with the same model returns an identical score for 96 percent of Claude's variable-pitch cells and 91 percent of GPT's. Scoring at temperature 0 is not deterministic either, so reproducibility is high but not perfect. Appendix Table B.9 reports run-to-run differences.

The index survives on the reproducible fraction of its own variables. Scores that move between

runs would matter only if the index leaned on them. Rebuilt by the same rule on the 83 variables whose scores reproduce on rerun, judgment reads +16.3 percentage points at interest and +7.6 at the handshake; on the 84 flagged stable or moderate against the baseline codebook, +16.2 and +7.2; and on the 18 flagged stable, +14.0 and +5.2. The full index is +15.4 and +8.0. An eighth of the variables carries most of it, and nothing in the result depends on the variables whose scores move between runs. Appendix Table B.10 has every subset.

The list does not return; a different one does the same job. The second question is whether the same list returns, and if not, whether a different one would have done the same job. Four regenerated runs invented their own lists from the training transcripts and scored every pitch against their own codebooks, never having seen the paper's. The lists do not match ours. Each run returns between 61 and 75 percent of the paper's variables under the same name, and between 66 and 82 percent once a variable measuring the same thing under a different name is counted as a match; each also proposes between 37 and 45 variables the paper does not have. The models do not agree with themselves on what to measure. They do agree on what it comes to: built into indices by the paper's rule, the regenerated sets correlate with the paper's index at 0.87 to 0.90 and reach the same place. Appendix B.1 lists every run and what differed between them.

The two providers are not interchangeable. Read with Claude's 45 variables alone the index carries +11.7 percentage points at interest and +2.4 at the handshake, the second not distinguishable from zero; with GPT's 116 alone, +15.2 and +8.6; pooled, +15.4 and +8.0. The same asymmetry appears in the regenerated sets. The interest result holds under both providers; the handshake result is carried primarily by GPT. Pooling both providers reduces the measure's dependence on a single provider rather than creating the association, and the pooled estimate is not distinguishable from the stronger provider's own. On the other hand, pooling further runs of one model adds little. Appendix Table B.10 has both providers reported separately.

Recognition does not account for the result. Both models were shown each anonymized transcript and asked to name the venture and whether they remember the outcome. A claimed memory of the outcome is worth little against the benchmark of always saying "deal". More to the point, the result does not need the recognized pitches. On the 214 neither model named, judgment reads +16.7 percentage points at interest and +14.7 at the handshake, both at $p < 0.001$ and both at least as large as on the full sample. Recognized ventures do score higher on judgment, by 0.27 standard deviations for Claude's recognition and 0.34 for GPT's, both at $p < 0.001$. But adding recognition to the main regression changes nothing. Judgment stays at +15.2 and +8.1 beside Claude's recognition and at +14.9 and +7.4 beside GPT's, against +15.4 and +8.0 without either, and recognition itself is not significant at five percent for either model at either decision. Appendix B.2 reports the probe and Appendix Table B.10 the estimates.

Measurement and index

All four indices are built by one rule, set out in Section 4.5: each variable is standardized, multiplied by a sign fitted on the *handshake*, and averaged. Two steps in that rule could carry an association on their own, so each is checked directly: i) the signs are read from the same outcome the index is then used on, and ii) the judgment variables were written from part of the sample.

Cross-fitting the signs does not manufacture the result. Every variable is given a sign fitted on the *handshake*. A sign read off all 642 pitches already knows which pitches got a deal, so the index built from it carries part of the answer. Cross-fitting gives each pitch a sign fitted on other pitches only. To show that the difference matters, we break the link between a pitch and its outcome: the deal and no-deal labels are randomly reassigned across the 642 pitches, keeping the same number of each, so nothing real is left to find. We then build the language index both ways against those scrambled labels, one hundred times. Signs read off all 642 pitches still find a median of +11.4 percentage points across those draws. Cross-fitted signs find +0.5. On the real *handshake* the two give +12.4 and +5.0. However, it is important to note that even with cross-fitting, the scrambled labels give a coefficient significant at five percent in 19 of the 100 draws, where a correct test would give only five. With signs read off the whole sample it is 100 of 100. Hence cross-fitting closes the distance between 100 and 19; what remains between 19 and five is the limitation of Section 4.5, that the standard errors treat the index as given rather than as something built from the same data. Appendix Table B.11 has the distribution under both constructions.

Variables chosen on the training pitches. The judgment variables were generated by the models reading the 449 training pitches, so those pitches are in sample for which variables exist and the 193 are held out from variable generation. Judgment reads +16.1 percentage points at interest ($p < 0.001$) and +8.9 at the handshake ($p < 0.001$) on the 449, against +14.2 ($p = 0.004$) and +7.3 ($p = 0.059$) on the 193, and in a pooled model on all 642 the index's interaction with the held-out indicator is −0.00 at interest ($p = 0.99$) and −0.10 at the handshake ($p = 0.63$). The stricter version of the same question freezes every fitted component on the 449, the preprocessing, every sign and the index's own scale, and applies it unchanged to the 193, so that no outcome among the held-out pitches enters anything. Judgment carries +12.0 percentage points at interest ($p = 0.006$) and +6.6 at the handshake ($p = 0.086$) under that construction, against +14.0 ($p = 0.004$) and +7.2 ($p = 0.059$) with the index built on all 642.

Empirical strategy

Section 5 fits one logit on the four indices with one control set, and reads the prediction exercise off fixed splits. Three things could carry those results on their own, so each is checked: i) the controls, the definition of each outcome and the weighting inside the indices; ii) the ensemble and the point at which the data preparation is fitted; and iii) the number of tests the paper reports.

Controls, outcome definitions and weighting. Everything in Table 3 is estimated four times: with the eleven structural controls of that table, with the full hand-coded set, with the season, the

founder counts and each seated investor, and with founder-group indicators and the season alone. In the joint model judgment's effect on interest ranges from +14.9 to +15.4 percentage points and on the handshake from +7.9 to +8.4; hard finance from +3.3 to +4.0 and +4.1 to +4.7; soft finance from +7.4 to +8.3 at *completion*; and language stays between +0.0 and +2.8 at every stage, with a smallest p of 0.38. The structural controls alone predict no better than chance out of sample, at AUCs of 0.489 and 0.451, so they are not carrying the comparison. Interest survives being defined differently: redefined as at least one offer, which 492 of the 642 pitches draw, judgment carries +10.8 ($p < 0.001$) and hard finance +3.7 ($p = 0.029$), with the other two insignificant. And the language null is not an artifact of the weighting: rebuilt with the 192 variables weighted one by one rather than by family, the joint-model coefficient moves by at most 0.8 of a percentage point at any stage. Appendix Table B.12 has all twelve coefficients under each control set.

The prediction ordering does not depend on the ensemble or on fitting the preparation inside each split. The seven learners were fixed in advance and the paper reads their soft vote, but each one separately tells the same story: at interest judgment beats the structural controls, language and hard finance in all seven, and at the handshake it beats the controls and language in all seven and hard finance in three, which is the same near-tie the ensemble reports. Fitting the capping, imputation, encoding and scaling once on the master split rather than inside every split moves every figure in Panel A of Table 11 by at most 0.003 of AUC and leaves the ordering intact. Other choices in the pipeline, the correlation floor, the collinearity threshold, the learners' hyperparameters, are not varied here. Appendix A.5 has the seven learners separately.

Multiple testing. The headline tests are the joint specification of each table: the twelve cells of Table 3 Panel B, the sixteen of Table 9 Panel B, and the full column of Table 8 for each of the two text indices. That is 30 tests, one rule, and it is the specification the discussion of each table rests on. Eleven are significant at five percent individually and nine survive the Benjamini-Hochberg procedure at a false discovery rate of five percent. The two findings the paper is built on clear it by a wide margin, judgment at interest at $q < 0.0001$ and at the handshake at $q = 0.0012$, as do judgment on operating status, on survival to year three and on the equity gap, and hard finance against a later funding round. Three clear it narrowly: judgment at a later funding round ($q = 0.035$), soft finance at *completion* ($q = 0.039$) and hard finance at the handshake ($q = 0.043$). Appendix Table B.13 lists all thirty.

7 Conclusion

Early-stage investment is negotiated in private, so we know more about what attracts investors to a venture than about what happens between first interest and money changing hands. This paper opens one such process to observation. On 642 *Shark Tank* pitches, we follow the same ventures from investor *interest* through the on-air *handshake* to post-show *completion*, we observe the *terms* on which each agreement was struck, and we follow the ventures afterwards. Each pitch is separated into four kinds of information: the realized figures the founder states, the forward claims stated as

numbers, how the founder speaks as a dictionary counts it, and a language model's judgment of the venture read from the whole pitch.

The steps are associated with different information. LLM judgment is what goes with drawing *interest*, whether the association is measured by a coefficient or by out-of-sample prediction, and it is also what goes with a better price: a standard deviation is worth 2.7 percentage points of the company at a given ask. At the *handshake*, judgment still leads, but its margin narrows and the stated figures begin to count beside it. At *completion*, judgment no longer predicts, and the only index that does is the founder's own forward claims, in the direction that the higher the numbers put on the future and on the terms, the more often the agreed deal fell through. Counted language adds nothing with the three decisions once judgment is in the regression, although both are soft information by the same definition.

That judgment is not only a mirror of investor taste is visible after the show. Ventures that read well are more often operating at the collection date, more often alive three years after airing and more often financed again, and these associations hold with the handshake itself in the regression.

All of this is association rather than causation. The judgment measure is a model score, the setting offers no exogenous variation, and a televised negotiation is not a partners' meeting: the sums are small, the median ask is $250,000 and the diligence is compressed. What the setting offers is a sequence of real decisions on real money for the same ventures, with the information available at each one measured the same way.

Two things seem worth carrying forward. A finding that some factor matters for early-stage funding may depend on which step it was measured at. And a handshake is not an investment: among the agreements whose outcome we can verify, 39.3 percent did not complete, and the information that predicts the agreement is not the information that predicts its survival.

References

American Broadcasting Company (2026). About Shark Tank. ABC.com. "28 Emmy nominations over 17 seasons"; accessed 29 September 2026.

Anderson, M. L. (2008). Multiple inference and gender differences in the effects of early intervention: A reevaluation of the Abecedarian, Perry Preschool, and Early Training projects. *Journal of the American Statistical Association 103*(484), 1481–1495.

Avery, M., E. Ip, A. Leibbrandt, and J. Vecci (2026). A brave new world of hiring: A natural field experiment on how asynchronous interviews and AI assessment reshape recruitment. CESifo Working Paper 12573, CESifo, Munich.

Barth, A., S. Mansouri, and F. Woebbeking (2023). "let me get back to you"—a machine learning approach to measuring nonanswers. *Management Science 69*(10), 6333–6348.

Bartlett, R., A. Morse, R. Stanton, and N. Wallace (2022). Consumer-lending discrimination in the fintech era. *Journal of Financial Economics 143*(1), 30–56.

Bem, S. L. (1974). The measurement of psychological androgyny. *Journal of Consulting and Clinical Psychology 42*(2), 155–162.

Berg, T., V. Burg, A. Gombović, and M. Puri (2020). On the rise of fintechs: Credit scoring using digital footprints. *The Review of Financial Studies 33*(7), 2845–2897.

Bernstein, S., A. Korteweg, and K. Laws (2017). Attracting early-stage investors: Evidence from a randomized field experiment. *Journal of Finance 72*(2), 509–538.

Boyd, R. L., A. Ashokkumar, S. Seraj, and J. W. Pennebaker (2022). The development and psychometric properties of LIWC-22. Technical report, University of Texas at Austin, Austin, TX.

Canal, E. (2016, October). We fact-checked seven seasons of Shark Tank deals. here are the results. *Forbes*. Accessed: 2026-08-19.

Cao, S., W. Jiang, J. Wang, and B. Yang (2024). From man vs. machine to man + machine: The art and AI of stock analyses. *Journal of Financial Economics 160*, 103910.

Chen, Y., B. T. Kelly, and D. Xiu (2022). Expected returns and large language models. Available at SSRN: 4416687.

Chernozhukov, V., D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, W. Newey, and J. Robins (2018). Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal 21*(1), C1–C68.

Chopra, F. and I. Haaland (2023). Conducting qualitative interviews with ai.

CNBC Make It (2018, May). Barbara Corcoran breaks down what a day filming 'Shark Tank' is really like. CNBC. Published May 3, 2018. Add byline. Accessed September 2026.

Cohen, L., Y. Lu, and Q. H. Nguyen (2026). Mimicking finance. NBER Working Paper 34849, National Bureau of Economic Research.

Coqueret, G., J. Llull, F. Oswald, C. Pérignon, C. Scheuch, and L. Vilhuber (2026). Randomness in large language models: What researchers need to know (and report). *SSRN Working Paper* (7191580). arXiv:2607.24372.

Crawford, V. P. and J. Sobel (1982). Strategic information transmission. *Econometrica 50*(6), 1431–1451.

Dalle, J.-M., M. den Besten, and C. Menon (2017). Using Crunchbase for economic and managerial research. OECD Science, Technology and Industry Working Papers 2017/08, Organisation for Economic Co-operation and Development, Paris.

Dawes, R. M. (1979). The robust beauty of improper linear models in decision making. *American Psychologist 34*(7), 571–582.

Fast, E., B. Chen, and M. S. Bernstein (2016). Empath: Understanding topic signals in large-scale text. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pp. 4647–4657.

Fiske, S. T., A. J. C. Cuddy, and P. Glick (2007). Universal dimensions of social cognition: warmth and competence. *Trends in Cognitive Sciences 11*(2), 77–83.

Fitzgerald, T. and K. Ray (2025, June). How do venture capitalists (actually) make decisions? internal evidence from a private startup accelerator. Working paper, Texas A&M University.

Fuster, A., P. Goldsmith-Pinkham, T. Ramadorai, and A. Walther (2022). Predictably unequal? the effects of machine learning on credit markets. *The Journal of Finance 77*(1), 5–47.

Gentzkow, M., B. Kelly, and M. Taddy (2019). Text as data. *Journal of Economic Literature 57*(3), 535–574.

Gompers, P., W. Gornall, S. N. Kaplan, and I. A. Strebulaev (2020). How do venture capitalists make decisions? *Journal of Financial Economics 135*(1), 169–190.

Gompers, P. A. (1995). Optimal investment, monitoring, and the staging of venture capital. *Journal of Finance 50*(5), 1461–1489.

González-Uribe, J., R. Klingler-Vidra, S. Wang, and X. Yin (2026). The broader role of venture capital due diligence. *The Review of Financial Studies 39*(7), 2018–2063.

Grossman, S. J. (1981). The informational role of warranties and private disclosure about product quality. *Journal of Law and Economics 24*(3), 461–483.

Hochberg, Y. V., A. Ljungqvist, and Y. Lu (2007). Whom you know matters: Venture capital networks and investment performance. *The journal of finance 62*(1), 251–301.

Horton, J. J., A. Filippas, and B. S. Manning (2023). Large language models as simulated economic agents: What can we learn from Homo Silicus? NBER Working Paper 31122, National Bureau of Economic Research.

Hu, A. and S. Ma (2025). Persuading investors: A video-based study. *The Journal of Finance 80*(5), 2639–2688.

Huang, X., Z. Ivković, J. X. Jiang, and I. Y. Wang (2023). Angel investment and first impressions. *Journal of Financial Economics 149*(2), 161–178.

Hutto, C. J. and E. Gilbert (2014). Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the International AAAI Conference on Web and Social Media*, Volume 8, pp. 216–225.

Jabarian, B. and L. Henkel (2026). Voice AI in firms: A natural field experiment on automated job interviews. *arXiv preprint arXiv:2607.28222*.

Jagtiani, J. and C. Lemieux (2019). The roles of alternative data and machine learning in fintech lending: Evidence from the lendingclub consumer platform. *Financial Management 48*(4), 1009–1029.

Jetter, M. and K. Stockley (2023). Gender match and negotiation: Evidence from angel investment on shark tank. *Empirical Economics 64*(4), 1947–1977.

Kaniel, R., Z. Lin, M. Pelger, and S. Van Nieuwerburgh (2023). Machine-learning the skill of mutual fund managers. *Journal of Financial Economics 150*(1), 94–138.

Kanze, D., L. Huang, M. A. Conley, and E. T. Higgins (2018). We ask men to win and women not to lose: Closing the gender gap in startup funding. *Academy of Management Journal 61*(2), 586–614.

Kaplan, S. N. and P. Strömberg (2001). Venture capitalists as principals: Contracting, screening, and monitoring. *American Economic Review 91*(2), 426–430.

Kaplan, S. N. and P. Strömberg (2004). Characteristics, contracts, and actions: Evidence from venture capitalist analyses. *The journal of finance 59*(5), 2177–2210.

Kim, A., M. Muhn, and V. Nikolaev (2024). Financial statement analysis with large language models. Chicago Booth Research Paper No. 2024-65; Fama-Miller Working Paper.

Kling, J. R., J. B. Liebman, and L. F. Katz (2007). Experimental analysis of neighborhood effects. *Econometrica 75*(1), 83–119.

Lavanchy, M., P. Reichert, and A. Joshi (2022). Blood in the water: An abductive approach to startup valuation on ABC's shark tank. *Journal of Business Venturing Insights 17*, e00305.

Lerner, J. and R. Nanda (2020). Venture capital's role in financing innovation: What we know and how much we still need to learn. *Journal of Economic Perspectives 34*(3), 237–261.

Li, F. (2008). Annual report readability, current earnings, and earnings persistence. *Journal of Accounting and Economics 45*(2–3), 221–247.

Liberti, J. M. and M. A. Petersen (2019). Information: Hard and soft. *Review of Corporate Finance Studies 8*(1), 1–41.

Lopez-Lira, A. and Y. Tang (2026). Can ChatGPT forecast stock price movements? return predictability and large language models. *Journal of Financial Economics 184*, 104335.

Loughran, T. and B. McDonald (2011). When is a liability not a liability? textual analysis, dictionaries, and 10-ks. *The Journal of Finance 66*(1), 35–65.

Loughran, T. and B. McDonald (2016). Textual analysis in accounting and finance: A survey. *Journal of Accounting Research 54*(4), 1187–1230.

Ludwig, J. and S. Mullainathan (2024). Machine learning as a tool for hypothesis generation. *The Quarterly Journal of Economics 139*(2), 751–827.

Lyonnet, V. and L. H. Stern (2024). Machine learning about venture capital choices. Technical Report 2022-03-002, Fisher College of Business.

Martens, M. L., J. E. Jennings, and P. D. Jennings (2007). Do the stories they tell get them the money they need? the role of entrepreneurial narratives in resource acquisition. *Academy of Management Journal 50*(5), 1107–1132.

Maxwell, A. L., S. A. Jeffrey, and M. Lévesque (2011). Business angel early stage decision making. *Journal of Business Venturing 26*(2), 212–225.

Milgrom, P. R. (1981). Good news and bad news: Representation theorems and applications. *The Bell Journal of Economics 12*(2), 380–391.

Mohammad, S. M. (2018). Word affect intensities. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan.

Mohammad, S. M. (2025). NRC VAD lexicon v2: Norms for valence, arousal, and dominance for over 55k english terms.

Mohammad, S. M. and P. D. Turney (2013). Crowdsourcing a word–emotion association lexicon. *Computational Intelligence 29*(3), 436–465.

Nicolas, G., X. Bai, and S. T. Fiske (2021). Comprehensive stereotype content dictionaries using a semi-automated method. *European Journal of Social Psychology 51*(1), 178–196.

Peduzzi, P., J. Concato, E. Kemper, T. R. Holford, and A. R. Feinstein (1996). A simulation study of the number of events per variable in logistic regression analysis. *Journal of clinical epidemiology 49*(12), 1373–1379.

Retterath, A. and R. Braun (2020). Benchmarking venture capital databases. *SSRN Electronic Journal*. SSRN Working Paper 3706108.

Short, J. C., J. C. Broberg, C. C. Cogliser, and K. H. Brigham (2010). Construct validation using computer-aided text analysis (CATA): An illustration using entrepreneurial orientation. *Organizational Research Methods 13*(2), 320–347.

Stein, J. C. (2002). Information production and capital allocation: Decentralized versus hierarchical firms. *Journal of Finance 57*(5), 1891–1921.

Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. *Journal of Finance 62*(3), 1139–1168.

The New York Times (2024, October). Want to understand america? watch Shark Tank. *The New York Times*. Interactive article. Accessed: 2026-08-06.

Wallace, D. (2025, December). 'Shark Tank' executive producer shares 7 facts you probably don't know (exclusive). Parade. Published December 10, 2025. Accessed September 2026.

A Additional results

A.1 Guest investors

Thirty of the 38 investors who appear on the panel over Seasons 1 to 16 are guests. They sit on the panel for one or more episodes alongside the regular investors and make offers and agreements on the same terms. Table A.1 reports their backgrounds and their investment activity outside the show, on the same basis as Table 2 in Section 2.

Table A.1: Background and Investment Activity of Guest Investors

			Crunchbase activity		Shark Tank
Investor	Entrepreneurial background	Business outcome	Investments	Exits	Sample deals
Emma Grede	Co-founded Good American and Skims	Operator; Skims valued in the billions	7	–	7
Alex Rodriguez	Former MLB player; co-owner, Minnesota Timberwolves	A-Rod Corp holdings	23	4	5
Rashaun Williams	Former Goldman Sachs banker; venture investor	Coinbase, Robinhood, Dropbox (IPOs)	20	6	5
Kendra Scott	Founded Kendra Scott (jewellery)	$1bn valuation (2016) after a minority sale to Berkshire Partners	5	1	4
Peter Jones	Founded Phones International; Dragons' Den investor	Wireless Logic, £38m (2011)	13	–	4
Chris Sacca	Founded Lowercase Capital	Twitter IPO, reported $1bn return	44	25	4
Todd Graves	Founded Raising Cane's	$5bn annual revenue (2025)	–	–	4
Matt Higgins	Co-founded RSE Ventures	Resy to American Express (2019)	4	5	3
Charles Barkley	NBA Hall of Famer and broadcaster	No venture vehicle reported	4	1	3
Rohan Oza	Built Vitaminwater at Glacéau	Glacéau to Coca-Cola, $4.2bn (2007)	13	7	3
Tony Xu	Co-founded DoorDash	DoorDash IPO (2020)	34	3	3
Kevin Hart	Founded Hartbeat (media company)	Operator	20	4	3
Nirav Tolia	Co-founded Nextdoor and Epinions	Nextdoor public listing (2021)	10	4	3
Gwyneth Paltrow	Founded Goop	Operator	18	4	2
Michael Rubin	Founded GSI Commerce and Fanatics	GSI to eBay, $2.4bn (2011)	13	–	2
Alli Webb	Co-founded Drybar	Drybar to Helen of Troy, $255m (2020)	2	–	2
Jamie Kern Lima	Founded IT Cosmetics	Sold to L'Oréal, $1.2bn (2016)	–	–	2
Jason Blum	Founded Blumhouse Productions	Operator	–	–	2
Jamie Siminoff	Founded Ring (smart doorbells)	Ring to Amazon, $1bn (2018)	4	2	2
Blake Mycoskie	Founded TOMS (footwear)	50% to Bain Capital (2014)	15	2	1
Bethenny Frankel	Founded Skinnygirl	Cocktail line to Beam, $120m (2011)	3	–	1
Sara Blakely	Founded Spanx	Majority to Blackstone at $1.2bn (2021)	5	3	1
Anne Wojcicki	Co-founded 23andMe	Public listing (2021); Chapter 11 and CEO exit (2025)	60	8	1

continued on the next page

Table A.1, continued

Investor	Entrepreneurial background	Business outcome	Crunchbase activity		Shark Tank
			Investments	Exits	Sample deals
Candace Nelson	Founded Sprinkles Cupcakes	Exited 2012	–	–	1
Steve Tisch	Film producer; co-owner, New York Giants	Escape Artists	–	–	1
Maria Sharapova	Tennis champion; founded Sugarpova	Therabody, Tonal, Supergoop!	8	–	1
Richard Branson	Founded the Virgin Group	Virgin Galactic listing (2019)	26	12	1
Ashton Kutcher	Co-founded A-Grade and Sound Ventures	Airbnb, Uber, Skype	78	31	0
Katrina Lake	Founded Stitch Fix	Stitch Fix IPO (2017)	13	2	0
Troy Carter	Music manager; founded Atom Factory	Uber, Spotify, Dropbox (IPOs)	27	15	0

Notes. The table reports the entrepreneurial backgrounds and investment activity of the guest investors, ordered by the number of agreements reached. *Investments* and *Exits* are recorded by Crunchbase as of 28 September 2026; an exit is a portfolio company that was acquired or went public. -- marks a figure Crunchbase does not record: four guest investors have no Crunchbase investor profile, and Jason Blum has a profile with no investment count recorded. For Michael Rubin the current profile reports no investment count and the figure is taken from the May 2025 Crunchbase bulk export. These measures are lower bounds, because Crunchbase records an investment only where a funding round is registered and therefore understates personal investments, particularly in consumer businesses. *Sample deals* counts the on-air agreements in which each investor took part, identified from the recorded agreement rather than from a hand-coded participation flag, so an agreement made jointly is counted once for each participating investor. Three guest investors sat on the panel without reaching an agreement. Summed across both tables the column gives 568 investor-agreement pairs over the 429 agreements, because an agreement made jointly appears once for each participant. Every agreement is subject to post-show due diligence. Entrepreneurial backgrounds and business outcomes were collected from company filings, press releases, trade press, and investors' published biographies in September 2026.

A.2 Coverage by season

Table A.2 sets what aired against what we observe, season by season. The aired totals are published counts, compiled pitch by pitch from the season pages cited in: Wikipedia season pages for Seasons 1 to 13, and Primetimer episode threads and *Shark Tank Blog* episode pages thereafter, where Wikipedia carries no pitch detail.

Table A.2: Aired pitches and sample coverage, by season

	Aired				Our sample		
Season	Episodes	Pitches	Deals	Rate (%)	Pitches	Coverage (%)	*Handshake* (%)
1	14	64	27	42.2	6	9.4	50.0
2	9	36	19	52.8	8	22.2	75.0
3	15	60	28	46.7	6	10.0	33.3
4	26	103	52	50.5	20	19.4	70.0
5	29	116	62	53.4	15	12.9	40.0
6	29	116	66	56.9	14	12.1	57.1
7	29	116	65	56.0	52	44.8	57.7
8	24	96	59	61.5	18	18.8	55.6
9	24	96	65	67.7	30	31.2	76.7
10	23	92	63	68.5	63	68.5	68.3
11	24	96	63	65.6	60	62.5	63.3
12	25	100	71	71.0	63	63.0	74.6
13	24	96	64	66.7	95	99.0	66.3
14	22	87	62	71.3	84	96.6	71.4
15	22	88	60	68.2	38	43.2	68.4
16	20	80	58	72.5	70	87.5	71.4
Total	359	1442	884	61.3	642	44.5	66.8

Notes. *Rate* is the share of aired pitches reaching a deal in the published counts; *Handshake* is the share reaching an agreement among the pitches we observe.

A.3 The terms of the agreement

Table A.3: Which part of LLM judgment moves the terms

Variable group	Variables	Terms	Decisions	
		Equity gap (points of equity) $N = 428$	*Interest* (pp) $N = 642$	*Handshake* (pp) $N = 642$
Traction and customers	33	-2.499^{**}	$+13.63^{***}$	$+5.71^{**}$
Business model and economics	16	-1.867^{***}	$+15.29^{***}$	$+9.46^{***}$
Go-to-market and distribution	19	-1.723^{**}	$+11.63^{***}$	$+5.49^{**}$
Operations and risk	18	-1.549^{*}	$+11.86^{***}$	$+9.74^{***}$
Team and founders	19	-0.698	$+7.57^{**}$	$+4.49^{*}$
Capital and deal	10	-0.579	$+8.85^{***}$	$+5.11^{**}$
Market and competition	9	-0.471	$+10.05^{***}$	$+3.67^{\dagger}$
Mission	6	-0.017	$+1.32$	$+1.58$
Product and defensibility	15	$+0.030$	$+8.45^{***}$	$+4.70^{**}$
The four groups above, bundled	86	-2.542^{***}	$+16.21^{***}$	$+9.40^{***}$
The other five, bundled	59	-0.672	$+13.09^{***}$	$+7.07^{***}$
All nine entered together: are they jointly zero?				
the four groups above		$F = 4.63$	$\chi^2 = 34.10$	$\chi^2 = 24.50$
		$p = 0.001$	$p < 0.001$	$p < 0.001$
the other five		$F = 0.19$	$\chi^2 = 19.22$	$\chi^2 = 4.32$
		$p = 0.965$	$p = 0.002$	$p = 0.505$
fit of that model		$R^2 = 0.176$	pseudo $R^2 = 0.104$	pseudo $R^2 = 0.062$

Notes. Each variable group is made into its own index by the same rule as the full one, and each row is one group entered on its own with the eleven structural controls and standard errors clustered on the episode. Variables: the number of distinct judgment variables in the group. The nine groups partition the 145 variables; 16 of them are scored by both models, and the two columns of such a variable are averaged before the group index is built, so no variable is counted twice.

The equity column is the result this table is for: least squares on the winsorized equity gap over the 428 agreements, holding the level of the ask, the equity offered and the valuation they imply fixed, in points of the company handed over. The *investor interest* and *handshake* columns are logit average marginal effects in percentage points of the probability of the decision, on all 642 pitches, without the ask held fixed. They are the same estimates as Table 4 reports and are reproduced here so that the terms can be read against the decisions group by group.

The bundle rows enter the groups named in the row as a single averaged index. The joint tests come from one model holding all nine groups at once: an F test for the equity column and a Wald χ^2 for the two logits, both on the clustered covariance. In that nine-at-once model only business model and economics is significant on its own at the five percent level on the equity gap, at about half the coefficient it carries alone, so the finding belongs to the bundle rather than to any one group.

$^{\dagger}\, p < 0.10$, $^{*}\, p < 0.05$, $^{**}\, p < 0.01$, $^{***}\, p < 0.001$.

A.4 Past the deal

A.4.1 Who Crunchbase covers

Section 6.7 takes every post-show outcome from the open-web collection and leaves Crunchbase out. Table A.4 gives the reason. It runs the same regression for each source: the outcome is one if that source established an outcome for the venture and zero if it did not, and the right-hand side is the structural observables the paper uses elsewhere.

Table A.4: Is coverage selected? Each source's coverage on venture characteristics

	Crunchbase	Open web			
	profile	Operating	Acquisition	Round	Alive y3
Covered	388 (60%)	593 (92%)	587 (91%)	371 (58%)	490 (92%)
Panel A. Each characteristic alone					
Implied valuation	$+13.8^{***}$	$+3.2^{**}$	$+0.9$	-5.5^{**}	$+4.4^{**}$
Ask amount	$+11.2^{***}$	$+2.4^{*}$	-0.1	-5.1^{**}	$+3.9^{**}$
Prior investment	$+10.8^{***}$	$+0.5$	$+0.5$	$-3.7^{\dagger}$	$+2.5$
Equity offered	-12.3^{***}	-2.8^{**}	$-1.5^{\dagger}$	$+4.6^{*}$	-3.3^{***}
Bootstrapped	-10.8^{***}	$+0.0$	-0.1	$+3.0$	-2.1
Profitable	-8.6^{***}	-1.1	$-2.3^{\dagger}$	$+1.7$	$-2.8^{\dagger}$
Current revenue	$+4.4^{*}$	$+1.6^{*}$	$+0.0$	-3.3	$+1.4^{\dagger}$
Expected revenue	$+5.8^{**}$	$+0.3$	$+0.6$	-1.6	$+0.7$
Solo founder	$-7.3^{\dagger}$	-1.4	-5.2^{*}	$+3.9$	$-4.3^{\dagger}$
Female-only founders	-4.1	$+1.0$	-3.2	$+4.3$	-1.8
Mixed-gender founders	-0.6	$+2.7$	$+7.5^{**}$	$+2.9$	$+5.1^{\dagger}$
Manufacturing	$+0.4$	$+2.4$	-2.1	-3.4	-2.5
Aired later (season)	$-3.8^{\dagger}$	$+2.5^{**}$	$+1.6^{\dagger}$	-7.1^{***}	$+1.1$
Investor interest	$+7.0^{\dagger}$	$+0.5$	$+0.5$	$+1.1$	$+1.3$
Handshake	$+2.6$	$+1.2$	$+0.5$	$+6.4^{\dagger}$	$+1.7$
Panel B. All characteristics together					
McFadden pseudo R^2	0.115	0.062	0.062	0.035	0.082
Joint χ^2 p	1.9×10^{-14}	0.13	0.082	0.0093	0.051
AUC, out of sample	0.696	0.570	0.568	0.539	0.488
Panel C. The four indices, added to all the characteristics					
Hard finance	$-3.7^{\dagger}$	$+3.3^{\dagger}$	$+0.5$	$+0.6$	$+2.9^{\dagger}$
Soft finance	$+3.9$	$+0.7$	$+1.2$	$+1.1$	$+3.6$
Language	$-3.7^{\dagger}$	-0.9	$+0.3$	$+4.1^{\dagger}$	-1.2
LLM judgment	$+6.6^{*}$	$+0.0$	-0.4	$+1.7$	$+1.5$

Notes. Logit average marginal effects in percentage points on the probability that the source established an outcome for that venture, over all 642 pitches, with standard errors clustered on the episode. Continuous characteristics are standardized, so each figure is the effect of a one-standard-deviation increase; the founder, industry and on-air characteristics are indicators. The AUC is from five-fold cross-validation, so it is out of sample; 0.5 is coverage unrelated to the characteristics. Panel C adds the four indices to the full set in Panel B. The alive-at-year-three column is estimated on the 534 pitches of Seasons 1–14, where the outcome is observable, and counts the two-pass agreement before the eight ventures whose survival is implied by their trading at the collection date. The round column is the open-web coverage of a named investor round, the outcome used in Table 9.
$^{\dagger}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

Crunchbase coverage is selected and open-web coverage is not. Appearing in Crunchbase is predicted by every financial characteristic that distinguishes a venture-capital deal from a small business: a standard deviation of the implied valuation raises it by 13.8 percentage points, the ask by 11.2 and prior investment by 10.8, while giving away more equity lowers it by 12.3, being bootstrapped by 10.8 and being profitable by 8.6. Together the characteristics call coverage with an out-of-sample AUC of 0.696, and LLM judgment adds 6.6 points on top of them. Crunchbase covers the ventures it exists to cover and misses the self-funded profitable business, which on this sample is the majority.

The open-web outcomes behave differently. Their coverage is close to unpredictable, with AUCs between 0.488

and 0.570, and no index predicts it at the five percent level; the largest cell in Panel C is language on the funding round, +4.1 points at ten percent. The funding round is the only one whose joint test rejects, and what drives it is when the pitch aired, −7.1 points per standard deviation of the season, with a smaller contribution from having reached a handshake, +6.4: a venture that struck a deal on air generates follow-up reporting on whether the money arrived. Panel C of Table 9 holds the *handshake* fixed for that reason.

A.4.2 The two outcomes as Crunchbase records them

Table A.5: Follow-on capital and acquisition as Crunchbase records them

	Follow-on capital ($N = 388$)	Acquisition ($N = 388$)
Panel A. One kind of information at a time		
Hard finance (H_{finance})	−2.6	+0.1
	(2.8)	(0.8)
Soft finance (S_{finance})	+5.5*	+0.6
	(2.7)	(1.0)
Language (S_{language})	−3.3	+0.6
	(2.8)	(1.2)
LLM judgment (S_{judgment})	−1.0	+0.2
	(2.5)	(1.6)
Panel B. All four together		
Hard finance (H_{finance})	−2.6	−0.2
	(3.1)	(1.0)
Soft finance (S_{finance})	+7.0*	+0.6
	(3.0)	(1.1)
Language (S_{language})	−3.2	+0.7
	(3.0)	(1.4)
LLM judgment (S_{judgment})	−1.6	−0.2
	(2.9)	(1.8)
Panel C. All four together, the handshake held fixed		
Hard finance (H_{finance})	−2.8	−0.3
	(3.1)	(1.0)
Soft finance (S_{finance})	+7.4*	+0.7
	(3.1)	(1.0)
Language (S_{language})	−3.4	+0.6
	(3.1)	(1.4)
LLM judgment (S_{judgment})	−2.1	−0.4
	(3.0)	(1.8)

Notes. Logit average marginal effects in percentage points per standard deviation of the index, standard errors beneath, clustered on the episode, over the 388 ventures with a matched Crunchbase profile. Follow-on capital is recorded for 138 of them (35.6 percent) and an acquisition for 18 (4.6 percent). The funding flag is one where Crunchbase records a round announced after the episode aired; it does not exclude the investment agreed on the show, so it is not the named investor round of Table 9. Follow-on capital carries the three post-show controls; acquisition carries none, since 18 events support none. Specifications otherwise match Table 9.
$^{\dagger}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

On Crunchbase variables LLM judgment says nothing: −1.0 percentage points on follow-on capital and +0.2 on acquisition, −2.1 and −0.4 in Panel C, against +4.5 and +2.5 on the open record. Hard finance keeps its negative sign on follow-on capital, −2.6 against the open record's −3.3, and is the index the two sources agree on. The acquisition column rests on 18 events with no controls and should not be read at all: every cell in it is under one percentage point with a standard error above it.

The funding column is a different variable from the funding column of Table 9, and the difference is not one of timing. Crunchbase's flag is dated, it is one where the venture has a round announced after the episode aired, for all 388 ventures, but it does not exclude the investment agreed on the show. Crunchbase's own round records make

this visible. Of the 237 post-airing rounds it itemizes, 31 carry an amount within five per cent of the sum agreed on air; 20 of those fall within a week of the episode and 18 are recorded as angel investments, concentrated in the later seasons, nine in Season 13 and six in Season 14. The timing says the same without the amounts: the first post-airing round arrives a median of 31 days after the episode for ventures whose deal closed and 242 days for ventures that left without a handshake. This column is therefore a measure rather than the funding round presented in Section 6.7. It does not account for what the indices do: judgment reads −1.6 percentage points on the flag among the 264 linked ventures whose on-air deal did not close, where no investor money could have entered the measure at all.

A.5 Out of sample prediction

Table A.6: Paired comparisons behind Table 11

	Investor interest			Handshake		
Comparison	Δ	Wins	p	Δ	Wins	p
Panel A. Against the structural controls						
Hard finance − baseline	+0.076	48	< 0.001	+0.125	50	< 0.001
Soft finance − baseline	+0.047	48	< 0.001	+0.020	38	< 0.001
Language − baseline	+0.058	48	< 0.001	+0.050	44	< 0.001
LLM judgment − baseline	+0.163	50	< 0.001	+0.120	50	< 0.001
All four together − baseline	+0.141	50	< 0.001	+0.136	50	< 0.001
Panel B. Every pair of information sets						
Hard finance − Soft finance	+0.029	39	< 0.001	+0.105	49	< 0.001
Hard finance − Language	+0.018	31	0.006	+0.075	48	< 0.001
Hard finance − LLM judgment	−0.087	0	< 0.001	+0.005	29	0.304
Hard finance − All four together	−0.065	2	< 0.001	−0.012	16	0.039
Soft finance − Language	−0.012	21	0.062	−0.030	14	< 0.001
Soft finance − LLM judgment	−0.116	0	< 0.001	−0.100	2	< 0.001
Soft finance − All four together	−0.095	0	< 0.001	−0.116	0	< 0.001
Language − LLM judgment	−0.105	0	< 0.001	−0.070	4	< 0.001
Language − All four together	−0.083	0	< 0.001	−0.087	0	< 0.001
LLM judgment − All four together	+0.022	42	< 0.001	−0.016	14	0.010

Notes. Δ is the mean difference in test AUC within a split, where both sides were scored on the same test pitches; *wins* is the number of the 50 splits in which the first is ahead; p is a Wilcoxon signed-rank test on the 50 paired differences. The splits are resamples of the same 642 pitches rather than independent samples, so the p-values describe how consistently one set is ahead across resamples rather than testing a population hypothesis, and the win count is the more informative column. Every information set carries the structural controls.

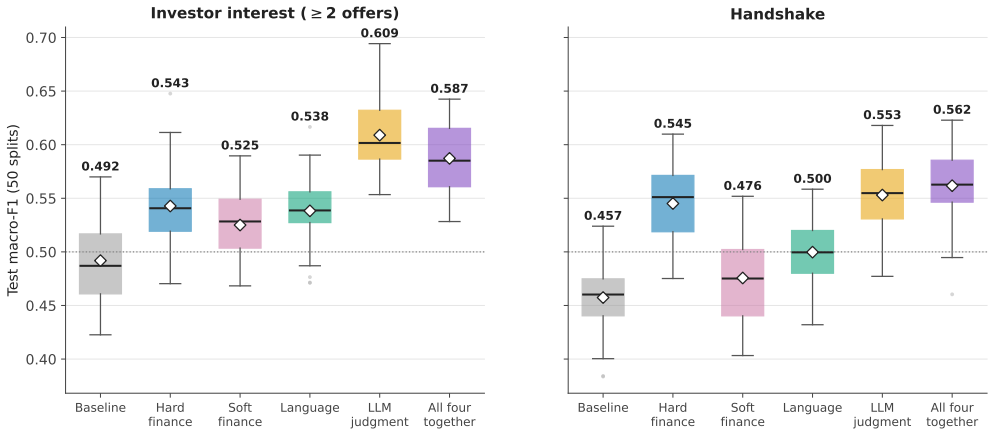


Figure A.1: Out-of-sample prediction by information set

Notes. Each box is the distribution of the Macro-F1 score of the seven-learner ensemble over 50 train/test splits; the diamond and the number mark the mean; the dotted line is 0.5. Each information set enters with the structural controls.

A.6 The direction of every variable

Every index in the paper takes its signs by cross-fitting on the handshake, so an index coefficient carries a magnitude but not a direction that can be read from its name. This table is where the directions are. Each of the 365 variables in the four sets is entered on its own, standardized, with the same eleven structural controls as Table 3, and no sign is applied to it anywhere. A positive number therefore means exactly what it says: pitches with more of that variable were more likely to reach the outcome.

Table A.7: The direction of every variable, entered one at a time

Variable	Group	Investor interest	Hand-shake	Comple-tion
Panel A. Financial figures (28)				
revenue_current	Hard finance	$+11.5^{***}$	$+3.6^{*}$	-2.8
is_profitable	Hard finance	$+7.6^{***}$	$+8.0^{***}$	$+3.2$
revenue_expected	Soft finance	$+5.5^{*}$	-1.2	-12.5^{***}
implied_valuation	Soft finance	$+5.2^{*}$	-3.1	-9.1^{**}
units_sold	Hard finance	$+4.3^{\dagger}$	$+3.2^{\dagger}$	$+0.3$
bootstrapped	Hard finance	$+3.2$	$+0.7$	$+3.1$
gross_margin	Hard finance	$+2.8$	$+3.5^{\dagger}$	$+0.7$
ask_amount	Soft finance	$+2.5$	-4.8^{*}	-10.8^{***}
development_expense	Hard finance	$+1.7$	$+0.0$	$+1.7$
fixed_costs_monthly	Hard finance	$+1.5$	$+2.6$	-2.3
unit_price	Hard finance	$+0.9$	-3.3	$-4.9^{\dagger}$
market_size_customers	Soft finance	$+0.7$	$+2.9$	$+1.9$
marketing_spend	Hard finance	$+0.4$	-4.1	-0.3
market_size_revenue	Soft finance	$+0.3$	$+0.5$	$-5.4^{\dagger}$
gross_margin_stated	Hard finance	$+0.3$	$+0.8$	$+1.1$
unit_cost	Hard finance	-0.2	-2.1	-1.8
revenue_growth_past	Hard finance	-0.2	-1.1	$+5.8$
personal_liability	Hard finance	-0.3	-0.6	$+7.1^{\dagger}$
customer_acquisition_cost	Hard finance	-0.4	-1.8	$+2.2$
has_marketing_spend	Hard finance	-0.8	-1.5	-2.1
payback_period	Soft finance	-0.8	-1.7	$+0.4$
has_debt	Hard finance	-1.0	-3.5^{*}	$+0.4$
founder_capital_invested	Hard finance	-1.4	-2.8	$+4.9$
nonfounder_equity_share	Hard finance	-1.5	$+0.2$	-3.9
prior_investment	Hard finance	-1.6	-0.5	-3.5
funding_raised	Hard finance	-2.2	-1.7	$+0.1$
debt_amount	Hard finance	-3.1	-2.8	$+4.8$
equity_ask	Soft finance	-7.3^{***}	-0.3	$+2.2$
Panel B. LLM judgment variables (145)				
pricing	Business model and economics	$+15.6^{***}$	$+10.4^{***}$	-2.0
product_market_fit	Traction and customers	$+15.3^{***}$	$+8.3^{***}$	-0.6
execution	Operations and risk	$+15.2^{***}$	$+8.9^{***}$	$+0.1$
customer_acquisition	Traction and customers	$+15.2^{***}$	$+7.5^{***}$	-1.4
unit_economics	Business model and economics	$+14.9^{***}$	$+11.2^{***}$	$+2.4$
customer_traction	Traction and customers	$+14.8^{***}$	$+7.4^{***}$	-0.1
cost_structure	Business model and economics	$+14.0^{***}$	$+11.1^{***}$	$+1.3$
product_quality	Product and defensibility	$+14.0^{***}$	$+9.3^{***}$	$+1.7$
customer_validation	Traction and customers	$+14.0^{***}$	$+6.7^{***}$	-1.4
business_model	Business model and economics	$+13.9^{***}$	$+9.1^{***}$	-0.6
growth	Traction and customers	$+13.8^{***}$	$+6.8^{***}$	$+1.6$

Table A.7, continued

Variable	Group	Investor interest	Hand-shake	Comple-tion
traction	Traction and customers	$+13.7^{***}$	$+6.6^{***}$	-0.0
go_to_market	Go-to-market and distribution	$+13.1^{***}$	$+6.6^{***}$	-2.4
scalability	Business model and economics	$+13.1^{***}$	$+6.2^{**}$	-0.8
exit_potential	Capital and deal	$+13.1^{***}$	$+3.3^{\dagger}$	-2.7
channel_fit	Go-to-market and distribution	$+12.0^{***}$	$+5.1^{**}$	-1.4
timing	Market and competition	$+11.8^{***}$	$+4.0^{*}$	-2.5
distribution	Go-to-market and distribution	$+11.0^{***}$	$+4.9^{**}$	-0.7
credibility	Team and founders	$+10.9^{***}$	$+7.2^{***}$	$+0.2$
focus	Operations and risk	$+10.7^{***}$	$+6.4^{***}$	$+0.8$
product_stage	Product and defensibility	$+10.1^{***}$	$+6.4^{***}$	$+3.9$
distribution_channel	Go-to-market and distribution	$+10.0^{***}$	$+3.6^{\dagger}$	-1.8
social_proof	Traction and customers	$+9.7^{***}$	$+4.0^{*}$	-1.8
channel_strategy	Go-to-market and distribution	$+9.6^{***}$	$+4.6^{*}$	$+1.6$
strategic_fit	Go-to-market and distribution	$+9.4^{***}$	$+6.0^{**}$	$+0.5$
sales_channel	Go-to-market and distribution	$+9.0^{***}$	$+3.9^{*}$	-1.9
team	Team and founders	$+8.7^{***}$	$+7.1^{**}$	-0.5
viral_traction	Traction and customers	$+8.6^{***}$	$+1.4$	$-5.8^{\dagger}$
supply_chain	Operations and risk	$+8.6^{***}$	$+7.1^{***}$	-2.0
brand_strength	Go-to-market and distribution	$+8.5^{***}$	$+3.7^{\dagger}$	-1.3
team_quality	Team and founders	$+8.4^{***}$	$+7.2^{***}$	-0.7
resilience	Operations and risk	$+8.4^{***}$	$+6.8^{***}$	-1.2
category	Market and competition	$+8.2^{***}$	$+2.3$	-0.3
sales_channels	Go-to-market and distribution	$+8.1^{***}$	$+1.9$	-0.9
differentiation	Product and defensibility	$+7.8^{***}$	$+3.5^{\dagger}$	$+2.2$
vision	Team and founders	$+7.7^{***}$	$+1.9$	$+1.3$
product_differentiation	Product and defensibility	$+7.7^{***}$	$+3.6^{*}$	$+2.2$
retailer_interest	Traction and customers	$+7.2^{***}$	$+3.3$	$+1.1$
channel_mix	Go-to-market and distribution	$+6.9^{**}$	$+2.0$	$+1.0$
moat	Product and defensibility	$+6.9^{***}$	$+3.8^{*}$	$+0.6$
growth_rate	Traction and customers	$+6.7^{**}$	$+1.3$	$+8.4^{**}$
use_of_funds	Capital and deal	$+6.6^{***}$	$+5.4^{**}$	$+3.7$
media_coverage	Traction and customers	$+6.3^{**}$	-0.3	$-6.2^{\dagger}$
team_composition	Team and founders	$+5.9^{*}$	$+5.4^{*}$	$+0.4$
retail_doors	Go-to-market and distribution	$+5.9^{**}$	$+0.6$	$+2.1$
customer_focus	Traction and customers	$+5.7^{*}$	$+5.1^{*}$	-0.4
market_size	Market and competition	$+5.7^{**}$	$+1.8$	-0.1
full_time_founders	Team and founders	$+5.6^{**}$	$+2.9$	-4.1
social_following	Traction and customers	$+5.6^{*}$	-1.1	-8.2^{*}
founder_fit	Team and founders	$+5.1^{**}$	$+3.1^{\dagger}$	$+1.5$
wholesale_mix	Go-to-market and distribution	$+5.0^{*}$	$+5.1^{*}$	$+3.9$
trade_secrecy	Product and defensibility	$+5.0^{*}$	$+3.7^{\dagger}$	$+2.5$

Table A.7, continued

Variable	Group	Investor interest	Hand-shake	Comple-tion
average_order	Business model and economics	+4.9*	+2.8	+0.4
inventory_status	Operations and risk	+4.9*	+5.0**	+2.1
market_growth	Market and competition	+4.7*	+0.7	−1.6
ownership	Capital and deal	+4.6*	+2.9	−1.5
full_time	Team and founders	+4.6*	+1.6	−2.6
branding	Go-to-market and distribution	+4.5*	+0.5	−3.9
prior_exit	Team and founders	+4.4*	−0.5	–
product_uniqueness	Product and defensibility	+4.4*	+1.7	−2.8
founder_experience	Team and founders	+4.3*	+2.2	−1.2
strategic_partnerships	Go-to-market and distribution	+3.9†	+2.0	+0.6
barriers_entry	Product and defensibility	+3.7*	+2.8	+1.1
return_on_ad_spend	Business model and economics	+3.6*	+2.8	−1.8
capacity	Operations and risk	+3.5†	+4.1*	+2.9
order_value	Business model and economics	+3.4†	+1.5	−2.4
founder_commitment	Team and founders	+3.3	+2.7	−2.9
ip_strength	Product and defensibility	+3.1†	+1.5	+0.8
domain_expertise	Team and founders	+3.1	+2.6	+0.7
partnerships	Go-to-market and distribution	+3.0	−0.2	+0.3
innovation	Product and defensibility	+3.0	+1.0	−0.3
commitment	Team and founders	+2.9	+1.8	−2.6
retail_presence	Go-to-market and distribution	+2.9	+1.8	+4.8
sales_efficiency	Business model and economics	+2.5	+1.8	−2.0
celebrity_endorsement	Traction and customers	+2.5	+2.3	+3.6
pre_sales	Traction and customers	+2.3	+0.1	+5.5*
founder_count	Team and founders	+2.2	−0.9	+1.1
has_cofounder	Team and founders	+2.2	+2.7	−2.0
conversion_rate	Traction and customers	+2.1	+0.6	−1.3
seasonality	Market and competition	+2.0	+3.5	+4.2
subscription_rate	Business model and economics	+1.7	+0.8	–
patent_status	Product and defensibility	+1.5	+1.0	+0.4
co_founder	Team and founders	+1.5	+3.2	−2.4
purchase_orders	Traction and customers	+1.5	+1.2	+6.8*
governance	Operations and risk	+1.2	+0.2	−4.0
patents	Product and defensibility	+1.2	+3.1	+0.6
outside_capital	Capital and deal	+0.9	−1.4	−1.2
inventory_risk	Operations and risk	+0.8	+1.4	−0.4
repeat_purchase	Traction and customers	+0.8	+1.2	+0.4
patent_protection	Product and defensibility	+0.7	+0.7	+1.2
pilot_results	Traction and customers	+0.5	−1.3	+3.4
manufacturing	Operations and risk	+0.2	+3.1	+6.2*
founder_invested	Team and founders	+0.1	+0.0	+0.3
customer_concentration	Traction and customers	−0.0	+2.6	+2.7
team_size	Team and founders	−0.0	−0.1	+7.3†
lead_time	Operations and risk	−0.1	+0.3	−5.4
retail_distribution	Go-to-market and distribution	−0.2	−3.2†	–

Table A.7, continued

Variable	Group	Investor interest	Hand-shake	Comple-tion
coachability	Team and founders	−0.3	+0.6	−5.2†
inventory_level	Operations and risk	−0.3	−0.9	+1.6
ownership_retained	Capital and deal	−0.4	+0.3	+5.0†
claim_risk	Operations and risk	−0.4	−3.3†	+4.7
inventory	Operations and risk	−0.6	−0.1	+0.9
franchising	Business model and economics	−0.6	+0.2	–
preorders	Traction and customers	−0.6	−2.3	+0.1
b2b_focus	Business model and economics	−0.8	−1.7	+0.3
subscription_model	Business model and economics	−0.8	−1.1	+1.4
sales_mix	Go-to-market and distribution	−1.1	+4.1†	−0.9
months_operating	Operations and risk	−1.4	−1.2	−2.7
repeat_rate	Traction and customers	−1.6	+0.6	+3.4
mission_alignment	Mission	−1.8	−2.7	+1.2
backlog	Traction and customers	−1.8	−1.5	−2.5
mission_strength	Mission	−1.8	−2.4	+2.1
cap_table	Capital and deal	−1.8	−1.8	−2.9
ip_risk	Product and defensibility	−1.9	−1.0	−4.9†
repeat_customers	Traction and customers	−2.0	−0.4	+2.8
licensing_deal	Business model and economics	−2.1	−0.8	+1.8
social_mission	Mission	−2.3	−1.1	+1.4
mission	Mission	−2.4	−3.0	+1.5
regulatory_risk	Operations and risk	−2.6	−3.2†	−2.8
mission_driven	Mission	−2.7	−1.7	+2.9
franchise_model	Business model and economics	−2.8	−1.6	–
competitive_intensity	Market and competition	−3.0	−1.6	−4.6†
licensing_model	Business model and economics	−3.1	−2.4	−2.7
eco_friendly	Mission	−3.1	+1.2	−0.3
competition	Market and competition	−3.2	−1.9	−4.4†
manufacturing_location	Operations and risk	−4.0*	+3.1	+7.6**
crowdfunding	Capital and deal	−4.0†	−4.2*	+1.3
liability_risk	Operations and risk	−4.4*	−5.5**	−8.0**
technology_risk	Product and defensibility	−5.5**	−5.9**	−6.4*
execution_risk	Operations and risk	−13.7***	−10.0***	−3.5
equity_offer	Capital and deal	–	–	–
runway	Capital and deal	–	–	–
valuation_gap	Capital and deal	–	–	–
partnership_pipeline	Go-to-market and distribution	–	–	–
category_growth	Market and competition	–	–	–
concentration	Market and competition	–	–	–
lifetime_value	Traction and customers	–	–	–
return_rate	Traction and customers	–	–	–
backorders	Traction and customers	–	–	–
churn	Traction and customers	–	–	–
ltv	Traction and customers	–	–	–
retention	Traction and customers	–	–	–
retention_rate	Traction and customers	–	–	–
sell_through	Traction and customers	–	–	–

Table A.7, continued

Variable	Group	Investor interest	Hand-shake	Comple-tion
waitlist	Traction and customers	–	–	–
Panel C. Language variables (192)				
quantity	Concreteness and numeracy	$+7.4^{***}$	+2.0	+1.2
number	Concreteness and numeracy	$+6.7^{**}$	+2.6	+2.4
millions	Concreteness and numeracy	$+5.5^{*}$	−0.3	-8.0^{**}
dollar_rate	Concreteness and numeracy	$+5.0^{*}$	$+3.3^{\dagger}$	+2.2
dollars	Concreteness and numeracy	$+4.6^{*}$	$+4.7^{*}$	−0.6
numbers	Concreteness and numeracy	$+4.5^{\dagger}$	+3.2	+1.1
empath_sadness_z	Tone and emotion	+4.4	$+4.3^{*}$	-6.1^{*}
top_word_share	Fluency and complexity	$+4.4^{*}$	+2.7	+0.5
fulfill	Entrepreneurial orientation	$+4.2^{*}$	$+3.6^{*}$	$+6.8^{*}$
empath_family_z	Self-presentation and address	+4.0	$+5.7^{*}$	−1.6
space	Concreteness and numeracy	$+3.3^{\dagger}$	−0.4	−2.0
certitude	Certainty and hedging	+3.2	+2.7	−1.4
empath_positive_emotion_z	Tone and emotion	+3.1	+2.4	−4.1
flesch	Fluency and complexity	+2.8	+1.6	−1.8
semantic_diversity	Fluency and complexity	+2.7	+2.7	−0.4
money	Concreteness and numeracy	+2.5	+0.2	+1.4
focuspast	Substance and time reference	+2.4	+0.0	−2.5
feeling	Tone and emotion	+2.4	+2.1	+1.0
lack	Tone and emotion	+2.3	+0.7	+1.3
lm_constraining_neg_per1k	Certainty and hedging	+2.2	−0.7	$-5.3^{\dagger}$
male	Self-presentation and address	+2.1	+0.5	+2.5
lm_litigious_neg_per1k	Certainty and hedging	+2.1	+0.1	–
SelfEfficacy	Entrepreneurial orientation	+2.1	−0.0	+0.6
eberhardt_ability_pos_per1k	Warmth and competence	+2.0	+3.3	$+5.6^{*}$
emoint_joy_mean	Tone and emotion	+1.9	+1.2	−0.1
emoint_disgust_mean	Tone and emotion	+1.8	+1.8	−3.6
polite	Warmth and competence	+1.8	+2.4	+3.8
percents	Concreteness and numeracy	+1.7	+1.8	−2.8
sentiment_mean	Tone and emotion	+1.6	+2.8	−0.8
empath_friends_z	Self-presentation and address	+1.4	+1.0	−1.2
vader_pos	Tone and emotion	+1.4	+2.4	−2.7
vader_pos_z	Tone and emotion	+1.4	+2.4	−2.7
bsri_femininity_pct	Warmth and competence	+1.4	−1.9	−0.0
emoint_anger_mean	Tone and emotion	+1.2	+2.2	+0.3
eberhardt_ability_neg_per1k	Warmth and competence	+1.2	−0.5	+4.8
Perception	Narrative and reasoning	+1.1	−1.7	−2.5
we	Self-presentation and address	+1.1	−1.2	+4.8
time	Concreteness and numeracy	+1.0	−1.7	−3.3
lm_strong_modal_neg_per1k	Certainty and hedging	+1.0	+2.8	+4.7
i	Self-presentation and address	+1.0	+1.5	−0.8
bsri_feminine_pos_per1k	Warmth and competence	+0.9	−2.0	−0.0
type_token_ratio	Fluency and complexity	+0.9	+0.8	+2.7
they	Self-presentation and address	+0.9	+2.3	−0.5
memory	Narrative and reasoning	+0.8	−2.2	+2.2
avg_sentence_length	Fluency and complexity	+0.8	+1.4	+2.4

Table A.7, continued

Variable	Group	Investor interest	Hand-shake	Comple-tion
emo_anger_per1k	Tone and emotion	+0.7	+1.4	−2.0
nrc_vad_coverage	Tone and emotion	+0.6	+2.5	−1.3
Risk_minisatation_transference	Certainty and hedging	+0.6	−0.4	−5.6†
Analytic	Fluency and complexity	+0.6	−0.3	+5.2†
Constraining	Certainty and hedging	+0.6	−0.7	−1.9
Creative	Entrepreneurial orientation	+0.5	−2.6	+2.6
last_sentence_sentiment	Tone and emotion	+0.5	+0.2	+0.4
pct_positive_sentences	Tone and emotion	+0.5	+0.8	−2.8
lm_uncertainty_pos_per1k	Certainty and hedging	+0.5	+0.3	−0.1
empath_negative_emotion_z	Tone and emotion	+0.4	−0.9	−2.1
emo_anger	Tone and emotion	+0.4	+1.9	+2.7
emoint_sadness_per1k	Tone and emotion	+0.4	−1.8	−5.0†
Risk	Entrepreneurial orientation	+0.3	−1.4	−1.8
emoint_fear_per1k	Tone and emotion	+0.3	−0.2	−5.4*
emo_fear_per1k	Tone and emotion	+0.3	+0.7	−4.4
emoint_sadness_mean	Tone and emotion	+0.3	+1.0	−2.5
tentat	Certainty and hedging	+0.3	+0.6	−3.5
emo_disgust_per1k	Tone and emotion	+0.3	+1.4	+0.2
Authentic	Self-presentation and address	+0.3	−2.0	−3.8
Tone	Tone and emotion	+0.2	−0.2	+0.2
emo_anx	Tone and emotion	+0.2	−1.2	−3.4
emotion	Tone and emotion	+0.2	−0.9	−3.2
Drives	Entrepreneurial orientation	+0.2	−0.9	+6.6*
emo_negative_per1k	Tone and emotion	+0.2	−1.2	−1.6
emoint_anger_per1k	Tone and emotion	+0.2	+0.8	−2.1
fatigue	Tone and emotion	+0.1	−0.3	−0.3
EntrepreneurialOrientation	Entrepreneurial orientation	+0.1	−2.1	−2.1
word_count	Fluency and complexity	+0.1	+1.4	−4.6
emoint_disgust_per1k	Tone and emotion	+0.1	+0.8	−0.3
lm_constraining_pos_per1k	Certainty and hedging	+0.0	−0.5	−0.4
lm_Uncertainty	Certainty and hedging	−0.0	−0.7	+0.1
acquire	Entrepreneurial orientation	−0.1	+0.2	−1.3
sad_warmth_neg_per1k	Warmth and competence	−0.1	−1.5	−2.7
emo_pos	Tone and emotion	−0.2	−1.8	−1.8
lm_strong_modal_pos_per1k	Certainty and hedging	−0.2	+3.0†	+0.5
sad_ability_neg_per1k	Warmth and competence	−0.2	−2.6	+3.8
StrongModal	Certainty and hedging	−0.2	+3.2†	+0.9
net_tone	Tone and emotion	−0.3	+2.3	+3.2
first_sentence_sentiment	Tone and emotion	−0.3	+2.8	−3.4
moral	Warmth and competence	−0.3	+1.7	−0.7
emo_sadness_per1k	Tone and emotion	−0.4	−1.3	−6.2*
Litigious	Certainty and hedging	−0.4	+1.3	+6.8*
vader_neu_z	Tone and emotion	−0.4	−1.6	+2.7
vader_neu	Tone and emotion	−0.4	−1.6	+2.7
Proactive	Entrepreneurial orientation	−0.5	−3.2†	−3.0
emoint_joy_per1k	Tone and emotion	−0.5	+1.8	−1.9
lm_weak_modal_pos_per1k	Certainty and hedging	−0.6	−2.4	−2.4
WeakModal	Certainty and hedging	−0.6	−2.6	−2.0
discrep	Certainty and hedging	−0.7	−0.7	−3.9
Clout	Warmth and competence	−0.7	−1.6	+1.0
nonflu	Fluency and complexity	−0.7	−0.2	−4.6†
allure	Warmth and competence	−0.8	−2.9	−3.1
questions	Fluency and complexity	−0.9	−2.0	−3.0
Reward	Entrepreneurial orientation	−0.9	−3.4†	−2.2

Table A.7, continued

Variable	Group	Investor interest	Hand-shake	Comple-tion
lm_negative_pos_per1k	Tone and emotion	−0.9	−1.1	−3.8
Uncertainty	Certainty and hedging	−1.0	−0.4	−2.0
Cognition	Narrative and reasoning	−1.0	−2.8	−5.3*
lm_litigious_pos_per1k	Certainty and hedging	−1.0	+1.4	+5.1†
need	Entrepreneurial orientation	−1.2	+0.9	−4.7†
emo_neg	Tone and emotion	−1.2	−1.0	−1.5
sad_warmth_pos_per1k	Warmth and competence	−1.2	−2.8	−1.0
want	Entrepreneurial orientation	−1.2	−2.3	+0.7
exclamations	Fluency and complexity	−1.2	−1.2	−4.7
n_sentences	Fluency and complexity	−1.2	+0.1	−5.6†
Mentor	Entrepreneurial orientation	−1.4	−2.1	+1.0
achieve	Entrepreneurial orientation	−1.5	−2.0	+4.1
comm	Warmth and competence	−1.5	−2.5	−3.1
lm_uncertainty_neg_per1k	Certainty and hedging	−1.5	−2.1	+0.4
visual	Narrative and reasoning	−1.5	−0.2	−3.0
prosocial	Warmth and competence	−1.5	+0.2	+1.7
Negative	Tone and emotion	−1.5	−1.1	−3.9
bsri_neutral_pos_per1k	Warmth and competence	−1.5	+1.9	−1.4
repeated_bigrams	Fluency and complexity	−1.6	−0.7	−5.0
lm_negative_neg_per1k	Tone and emotion	−1.6	−0.3	−2.9
motion	Narrative and reasoning	−1.7	−2.6	−0.6
nrc_valence_mean	Tone and emotion	−1.7	−0.8	+2.7
focusfuture	Substance and time reference	−1.7	−1.8	+3.6
Perservance_Tenacity	Entrepreneurial orientation	−1.8	−2.4	−3.1
lm_positive_pos_per1k	Tone and emotion	−1.8	+1.5	+2.3
contractions	Fluency and complexity	−1.8	−0.8	−5.2†
Innovative	Entrepreneurial orientation	−1.8	−2.7	+0.7
vader_neg_z	Tone and emotion	−1.8	−1.1	−0.7
vader_neg	Tone and emotion	−1.8	−1.1	−0.7
Affect	Tone and emotion	−1.8	−0.5	−3.2
negate	Certainty and hedging	−1.9	−3.1	−3.8
Optimistic	Tone and emotion	−2.0	−4.5*	−2.3
Stress_avg	Substance and time reference	−2.0	−2.3	−3.5
work	Substance and time reference	−2.1	−2.9	+4.3†
Guerrilla_skills	Entrepreneurial orientation	−2.3	−1.5	−2.8
sentiment_sd	Tone and emotion	−2.3	+0.7	−2.2
emo_joy_per1k	Tone and emotion	−2.4	+0.0	−0.8
emoint_trust_per1k	Tone and emotion	−2.4	+0.0	+1.5
emo_positive_per1k	Tone and emotion	−2.5	+0.0	+2.1
nrc_dominance_mean	Tone and emotion	−2.5	−3.0	+3.2
BigWords	Fluency and complexity	−2.5	−2.4	+0.8
conflict	Tone and emotion	−2.6	−0.2	+2.0
avg_word_length	Fluency and complexity	−2.6	−2.7	+1.3
emoint_trust_mean	Tone and emotion	−2.6	−2.1	−0.6
Positive	Tone and emotion	−2.7	+0.9	+3.5
tone_neg	Tone and emotion	−2.8	−1.2	−4.3
female	Self-presentation and address	−3.0	+2.9	+1.1
bsri_masculinity_pct	Warmth and competence	−3.0	+0.8	+5.1
sad_warmth_directed_per1k	Warmth and competence	−3.0	−1.4	+1.7
sad_ability_directed_per1k	Warmth and competence	−3.0	−3.5†	+5.8†
long_word_count	Fluency and complexity	−3.1	−3.0†	−0.0

Table A.7, continued

Variable	Group	Investor interest	Hand-shake	Comple-tion
emoint_anticipation_per1k	Tone and emotion	$-3.2^{\dagger}$	$+0.3$	-0.7
emo_trust_per1k	Tone and emotion	-3.3	$+0.4$	$+4.1$
emo_anticipation_per1k	Tone and emotion	$-3.3^{\dagger}$	$+0.2$	-0.7
emoint_fear_mean	Tone and emotion	$-3.4^{\dagger}$	$+0.4$	$+0.7$
curiosity	Entrepreneurial orientation	$-3.4^{\dagger}$	-1.9	$+1.9$
sad_ability_pos_per1k	Warmth and competence	$-3.4^{\dagger}$	-2.7	$+2.4$
syllables_per_word	Fluency and complexity	-3.4	-2.3	$+1.3$
auditory	Narrative and reasoning	$-3.5^{\dagger}$	-2.4	-2.3
nrc_arousal_high_per1k	Tone and emotion	$-3.5^{\dagger}$	-3.8^{*}	-0.1
cogproc	Narrative and reasoning	$-3.6^{\dagger}$	-2.1	-5.9^{*}
cause	Narrative and reasoning	$-3.6^{\dagger}$	-0.3	-0.9
emoint_surprise_mean	Tone and emotion	$-3.7^{\dagger}$	-1.5	-2.4
empath_communication_z	Self-presentation and address	$-3.8^{\dagger}$	-6.2^{**}	$-5.4^{\dagger}$
bsri_masculine_pos_per1k	Warmth and competence	$-3.9^{\dagger}$	-0.3	$+6.0^{\dagger}$
pct_negative_sentences	Tone and emotion	$-3.9^{\dagger}$	-0.9	-1.5
differ	Narrative and reasoning	-4.1^{*}	-2.9	-3.5
emo_surprise_per1k	Tone and emotion	-4.2^{*}	-0.9	$+0.9$
nrc_arousal_mean	Tone and emotion	-4.2^{*}	-3.0	$+2.6$
Social	Warmth and competence	-4.3^{*}	-2.4	$+0.5$
attention	Warmth and competence	-4.3^{*}	-4.9^{**}	-3.7
power	Entrepreneurial orientation	-4.4^{*}	-1.6	$+3.5$
lm_positive_neg_per1k	Tone and emotion	-4.5^{*}	-2.7	$+3.6$
you	Self-presentation and address	-4.7^{*}	-1.7	-3.9
emoint_surprise_per1k	Tone and emotion	-4.9^{*}	-1.4	$+0.9$
socbehav	Warmth and competence	-5.1^{*}	$-3.2^{\dagger}$	-0.3
focuspresent	Substance and time reference	-5.2^{*}	-4.3^{*}	-1.4
eberhardt_grindstone_pos_per1k	Warmth and competence	-5.8^{**}	-4.1^{*}	$+2.0$
lm_weak_modal_neg_per1k	Certainty and hedging	–	–	–
Opportunity	Entrepreneurial orientation	–	–	–
mental	Narrative and reasoning	–	–	$+1.1$
vader_compound	Tone and emotion	–	–	–
sentential_volatility	Tone and emotion	–	–	–
vader_compound_z	Tone and emotion	–	–	–
emoint_anticipation_mean	Tone and emotion	–	–	-6.3^{*}
empath_anger_z	Tone and emotion	–	–	-4.1
empath_fear_z	Tone and emotion	–	–	$+0.3$
emo_sad	Tone and emotion	–	–	-0.2
swear	Tone and emotion	–	–	$+10.0^{*}$
bsri_masculine_neg_per1k	Warmth and competence	–	–	–
bsri_feminine_neg_per1k	Warmth and competence	–	–	–
bsri_neutral_neg_per1k	Warmth and competence	–	–	–
eberhardt_grindstone_neg_per1k	Warmth and competence	–	–	–

Notes. Logit. Each variable enters on its own, standardized, with the eleven structural controls of Table 3; no sign is applied to any variable anywhere in this table, so a positive number means pitches with more of that variable were more likely to reach the outcome. Cells are changes in the probability of the outcome in percentage points for a one standard deviation rise in the variable, averaged over the pitches in the column. Standard errors are clustered on the episode. 642 pitches at investor interest and at the handshake, 313 at completion. A dash marks a variable too concentrated at a stage to be estimated there: fewer than ten pitches off its modal value, or a single pitch more than ten standard deviations from the mean. † $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

B Robustness Checks

B.1 Reproducibility of the judgment measure

The judgment measure was built in two stages: the models proposed variables and wrote a codebook, then scored every transcript against it. Both stages were repeated. Table B.8 lists every run. The 25 May 2026 run is the baseline and its codebook is the paper's codebook; the 15 September pass rescored every transcript against that same codebook with company and product names removed, and it is the pass the analysis uses. The remaining runs regenerated the variable list from the training transcripts and scored against their own codebooks, twice at temperature 0 and twice at 0.7, so that a rerun at a fixed temperature and a fixed codebook can be separated from a change of codebook and from a change of temperature.

Table B.8: Generation and scoring runs

Model	Run	Temperature	Task and codebook	Variables
Claude	25.05.2026	0 / 0.7	generation and scoring, names shown; baseline codebook	72
Claude	01.08.2026	0 / 0.7	generation and scoring, names shown; own codebook	91
Claude	09.08.2026	0 / 0.7	generation and scoring, names shown; own codebook	87
Claude	15.09.2026	0.7	scoring, names removed; baseline codebook *(the analysis)*	72
Claude	22.09.2026	0	generation and scoring, names removed; own codebook	89
Claude	23.09.2026	0.7	generation and scoring, names removed; own codebook	97
Claude	25.09.2026	0	rescoring, the 22.09 codebook reused	89
Claude	25.09.2026	0.7	rescoring, the 23.09 codebook reused	97
GPT	25.05.2026	default	generation and scoring, names shown; baseline codebook	148
GPT	01.08.2026	default	generation and scoring, names shown; own codebook	140
GPT	09.08.2026	default	generation and scoring, names shown; own codebook	123
GPT	15.09.2026	default	scoring, names removed; baseline codebook *(the analysis)*	148
GPT	22.09.2026	0	generation and scoring, names removed; own codebook	140
GPT	23.09.2026	0.7	generation and scoring, names removed; own codebook	135
GPT	25.09.2026	0	rescoring, the 22.09 codebook reused	140
GPT	25.09.2026	0.7	rescoring, the 23.09 codebook reused	135

Notes. "default" is the provider's own default temperature, which is 1 for GPT. The 25 May runs set temperature 0 for generation and 0.7 for scoring. Every run scored the same 643 transcript rows; one repeats a pitch already captured, so the analysis sample is 642 pitches.

Does the variable list come back? Of Claude's 72 baseline variables, 41 reappear under exactly the same name in *every* independently generated list (56.9 percent) and 49 reappear by meaning (68.1 percent); of GPT's 148, the counts are 66 (44.6 percent) and 73 (49.3 percent). Any single regenerated list returns more than that, between 61 and 75 percent of the baseline variables, because the variables that return are not the same ones each time. Pairwise, two independently generated lists of the same model overlap at a Jaccard index between 0.42 and 0.54 for Claude and between 0.49 and 0.57 for GPT. The variable list is one draw from a distribution, not a fixed instrument, which is why Section 6.9 reports the index rebuilt on the stable part of the list and rebuilt from independently generated lists altogether.

Do the scores come back? Table B.9 separates four kinds of change. A rerun with the codebook and the temperature held fixed is the clean measure of run-to-run noise and leaves 96 percent of Claude's scores and 91 percent of GPT's unchanged. Removing the company and product names changes 6 and 11 percent. Regenerating the codebook changes 20 and 24 percent, which is the order of magnitude of the variable-list instability above. Changing the temperature changes 19 and 26 percent, but those comparisons also change the codebook, so the clean statement about temperature is the first row read across: a rerun at temperature 0 changes 3.5 percent of Claude's scores and 8.4 percent of GPT's, against 4.7 and 8.9 percent at 0.7. Temperature 0 is not deterministic.

Table B.9: How often a score changes between runs

Model	What differs between the two runs	Comparisons	ρ	Identical
Claude	nothing (rerun, same codebook and temperature)	2	0.94	0.959
Claude	names shown rather than removed	1	0.91	0.940
Claude	the codebook	5	0.76	0.799
Claude	the temperature (and the codebook)	2	0.77	0.811
GPT	nothing (rerun, same codebook and temperature)	2	0.89	0.914
GPT	names shown rather than removed	1	0.83	0.890
GPT	the codebook	3	0.72	0.760
GPT	the temperature (and the codebook)	2	0.70	0.737

Notes. Averages over the variables scored in both runs of a comparison, equally weighted. ρ is the Spearman correlation between the two runs' scores; *Identical* is the share of pitch-variable cells with exactly the same score. Comparisons is the number of run pairs of that kind. The two rerun pairs are temperature 0 against temperature 0 and 0.7 against 0.7, each with the codebook held fixed; taken separately they leave 0.965 and 0.953 of Claude's scores identical and 0.916 and 0.912 of GPT's.

B.2 The recognition probe

Both models were shown each cleaned, anonymized transcript on its own and asked two questions: name the venture, and say whether you remember how the pitch ended. Neither question was part of the scoring, and the probe was run after it.

Claude names the venture for 293 of the 642 pitches (45.6 percent) and GPT for 382 (59.5 percent). Both recognize 247, one of the two recognizes 181, and neither recognizes 214; the two agree on 71.8 percent of pitches at $\kappa = 0.445$. A claimed memory of the outcome is close to uninformative. Claude claims one for 190 pitches, of which 58.4 percent did reach a handshake, and its recall is right for 63.7 percent of them, 5.3 points above always answering "deal" ($p = 0.080$, $\kappa = 0.33$). GPT claims a memory for 594 pitches and is right for 69.0 percent, 2.2 points above the same benchmark ($p = 0.138$, $\kappa = 0.29$). Both models' "deal" calls are more accurate than their "no deal" calls, 92.0 and 76.3 percent against 53.6 and 53.4, which is what guessing the base rate looks like.

B.3 The judgment index rebuilt

Table B.10 rebuilds the judgment index by the paper's own rule on narrower sets of its variables and re-estimates the joint model each time. Between any row and the first, the only thing that changes is which judgment variables are in the index.

Table B.10: The LLM judgment index rebuilt

How the index is built	Variables	*Investor interest*	*Handshake*
As built, all 145 variables	145	$+15.4^{***}$	$+8.0^{***}$
Rebuilt on the variables whose scores reproduce			
Scores identical on rerun	83	$+16.3^{***}$	$+7.6^{***}$
Stable or moderate against the codebook	84	$+16.2^{***}$	$+7.2^{***}$
Stable against the codebook	18	$+14.0^{***}$	$+5.2^{*}$
Rebuilt on one provider's variables			
Claude only	45	$+11.7^{***}$	$+2.4$
GPT only	116	$+15.2^{***}$	$+8.6^{***}$
Recognition			
Pitches neither model named	145	$+16.7^{***}$	$+14.7^{***}$
Holding Claude's recognition fixed	145	$+15.2^{***}$	$+8.1^{***}$
Holding GPT's recognition fixed	145	$+14.9^{***}$	$+7.4^{***}$

Notes. Logit average marginal effects in percentage points per standard deviation of the judgment index, with the other three indices and the eleven structural controls in every column and standard errors clustered on the episode. 642 pitches except where the row says otherwise: the row on the pitches neither model named reads 214 of them. *Variables* is how many judgment variables that row's index is built on, out of the 145 the paper uses. $^{\dagger}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

B.4 The handshake reshuffled

Every sign is fitted on the handshake, so an index signed on the same pitches it is then read on would carry part of the outcome by construction. Table B.11 reshuffles the handshake one hundred times and rebuilds the language index both ways against the reshuffled outcome.

Table B.11: The handshake reshuffled one hundred times

The language index, oriented	Signs read off every pitch	Signs cross-fitted
The reshuffled handshake, one hundred draws		
Median	$+11.4$	$+0.5$
5th to 95th percentile	$+9.4$ to $+14.1$	-4.7 to $+4.5$
Smallest and largest draw	$+8.3$ to $+16.0$	-9.4 to $+8.0$
Draws significant at five percent	100 of 100	19 of 100
Median out-of-sample AUC	0.644	0.505
The handshake as observed		
Average marginal effect	$+12.4^{***}$	$+5.0^{**}$
Out-of-sample AUC	0.656	0.562

Notes. The handshake is reshuffled across the 642 pitches so that no signal remains, and the language index is rebuilt against the reshuffled outcome one hundred times under each construction. Entries are logit average marginal effects in percentage points per standard deviation of the language index, with the eleven structural controls and standard errors clustered on the episode; the AUC is the out-of-sample area under the curve from the paper's own splits. Signs read off every pitch recover a large effect from an outcome that carries none; the cross-fitted construction centers on nothing. The rejection count is not a size check: the clustered standard error treats the index as a fixed regressor, as Section 4.5 states. $^{\dagger}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

B.5 Controls, outcome definitions and weighting

Table B.12 repeats the joint model of Table 3 Panel B under four control sets, under a wider definition of investor interest, and with the language index weighted variable by variable rather than by family.

Table B.12: The four indices under alternative specifications

	Investor interest	*Handshake*	*Completion*
Panel A. Four control sets, all four indices in the regression			
Estimated with the eleven structural controls of Table 3			
Hard finance	$+4.0^{*}$	$+4.7^{*}$	-2.2
Soft finance	$+0.7$	$+3.8^{*}$	$+8.3^{*}$
Language	$+0.5$	$+1.3$	$+2.7$
LLM judgment	$+15.4^{***}$	$+8.0^{***}$	$+2.4$
Estimated with the full hand-coded set, 53 columns			
Hard finance	$+3.7^{\dagger}$	$+4.1^{\dagger}$	-0.7
Soft finance	$+0.6$	$+4.4^{*}$	$+7.4^{*}$
Language	$+1.3$	$+1.1$	$+2.8$
LLM judgment	$+15.1^{***}$	$+8.4^{***}$	$+2.1$
Estimated with season, founders and the investors seated, 12 columns			
Hard finance	$+3.3^{\dagger}$	$+4.1^{*}$	-1.2
Soft finance	$+0.4$	$+3.8^{\dagger}$	$+8.0^{*}$
Language	$+0.0$	$+0.6$	$+1.0$
LLM judgment	$+15.4^{***}$	$+8.3^{***}$	$+1.5$
Estimated with founder group and season only, 3 columns			
Hard finance	$+3.7^{\dagger}$	$+4.5^{*}$	-2.5
Soft finance	$+0.2$	$+3.7^{*}$	$+8.2^{*}$
Language	$+0.5$	$+0.8$	$+1.8$
LLM judgment	$+14.9^{***}$	$+7.9^{***}$	$+1.8$
Panel B. Investor interest as at least one offer (492 of 642 pitches)			
Hard finance	$+3.7^{*}$	–	–
Soft finance	$+2.1$	–	–
Language	$+2.4$	–	–
LLM judgment	$+10.8^{***}$	–	–
Panel C. Language weighted variable by variable rather than by family			
Language, variable by variable	$+0.8$	$+0.8$	$+3.4$
Language, by family (the paper)	$+0.5$	$+1.3$	$+2.7$

Notes. Logit average marginal effects in percentage points per standard deviation of each index, all four entered jointly, standard errors clustered on the episode. Panel A repeats Table 3 Panel B under four control sets; the first block is the specification the paper uses and the other three are the alternatives. Panel B widens *investor interest* from two competing offers to at least one. Panel C rebuilds the language index weighting its 192 variables equally rather than its nine families equally. 642 pitches at *investor interest* and the *handshake*, 313 at *completion*. $^{\dagger}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

B.6 The thirty tests together

Table B.13 lists every headline test with its individual p and its Benjamini–Hochberg adjusted q.

Table B.13: The thirty headline tests, corrected together

Table	Index	Outcome	p	q
Table 3	LLM judgment	investor interest	0.0000	0.0000✓
Table 9	Hard finance	new funding round	0.0000	0.0001✓
Table 9	LLM judgment	alive at year three	0.0001	0.0007✓
Table 3	LLM judgment	handshake	0.0002	0.0012✓
Table 9	LLM judgment	operating at collection	0.0002	0.0012✓
Table 8	LLM judgment	equity gap	0.0012	0.0058✓
Table 9	LLM judgment	new funding round	0.0082	0.0353✓
Table 3	Soft finance	completion	0.0105	0.0393✓
Table 3	Hard finance	handshake	0.0129	0.0429✓
Table 3	Hard finance	investor interest	0.0362	0.1085
Table 3	Soft finance	handshake	0.0475	0.1297
Table 9	LLM judgment	acquisition	0.0615	0.1419
Table 8	Language	equity gap	0.0572	0.1419
Table 9	Soft finance	new funding round	0.0725	0.1450
Table 9	Hard finance	alive at year three	0.0721	0.1450
Table 9	Language	operating at collection	0.2059	0.3634
Table 9	Language	alive at year three	0.1984	0.3634
Table 9	Soft finance	acquisition	0.3427	0.5712
Table 3	Language	completion	0.3825	0.5827
Table 9	Hard finance	operating at collection	0.3884	0.5827
Table 9	Soft finance	alive at year three	0.4265	0.6092
Table 3	LLM judgment	completion	0.4694	0.6245
Table 3	Hard finance	completion	0.4788	0.6245
Table 3	Language	handshake	0.5494	0.6867
Table 9	Hard finance	acquisition	0.5801	0.6961
Table 9	Soft finance	operating at collection	0.6621	0.7640
Table 3	Soft finance	investor interest	0.7011	0.7790
Table 9	Language	acquisition	0.7348	0.7873
Table 9	Language	new funding round	0.7804	0.8073
Table 3	Language	investor interest	0.8279	0.8279

Notes. The headline tests are the joint specification of each table: Table 3 Panel B, Table 9 Panel B, and the full column of Table 8 for each text index. q is the Benjamini-Hochberg adjusted p-value over all thirty. A check marks the nine that survive at a false discovery rate of five percent.

C Prompts and Procedures

C.1 Step-by-step Procedure for *Shark Tank* Pitch Transcripts

Four research assistants transcribed each pitch using the Upmeet platform (an online AI tool for transcription) and manually checked each transcript. Every line was then assigned to one of ten categories in a separate classification step. Below is the step-by-step procedure for creating the transcripts.

- First, each research assistant is assigned to a specific season.
- Find the video: search for a YouTube video featuring a *Shark Tank* pitch. Most of the videos are found in the channel named *Shark Tank Global*. If a pitch is not available there, search YouTube using the company name or founder's name. This information is listed for each episode and pitch on IMDb.
- Use the Upmeet.ai platform
  - Copy the YouTube link and paste it into Upmeet.
  - We provide a standardized prompt to generate each transcript so that all transcripts follow the same format (e.g., Do not include timestamps. Format each line as: [Category] Speaker: Dialogue).
  - Upmeet produces transcripts in which every line is attributed to a speaker and carries a preliminary topic label. These preliminary labels are replaced in the categorization step below.
- Review and correct the transcript
  - Watch the video in segments and verify the transcript.
  - Correct speaker names: record the full name of the speaker.
- Categorize each line of the transcript
  - After the manual review, we discard the preliminary Upmeet labels and classify every line again with OpenAI's GPT-5.4 through the API, using one fixed prompt for all 642 pitches. The full prompt is reproduced at the end of this section.
  - The prompt asks the model to act as an annotator trained in venture capital decision-making and to assign each line to exactly one of ten categories.
    - `PRODUCT_TECH`: what the product or service is and how it works.
    - `OPERATIONS_EXECUTION`: how the product is produced and delivered.
    - `COMMERCIAL_STRATEGY`: how the company reaches customers and generates sales.
    - `FOUNDER_TEAM`: background and characteristics of the entrepreneurs.
    - `FINANCIALS_VALUATION`: financial performance and financial structure, including the initial ask.
    - `INVESTOR_VALUE_ADD`: non-monetary contributions of investors outside active bargaining.
    - `MARKET_COMPETITION`: external market environment and opportunity.
    - `NEGOTIATION`: active deal-making and bargaining.
    - `FINAL_COMMENTS`: final outcome and closing remarks.
    - `NON_SUBSTANTIVE`: fallback category for lines without business content.
- Prepare summary data in an Excel sheet for some basic analysis at the founder, investor, industry, and deal level.
- Provide a unique ID for each pitch (based on season, episode, and pitch) so that the pitch transcript and the Excel sheet can be connected.

Transcript Categorization Prompt: GPT

Model Name: `gpt-5.4`
Seed: 42
Temperature: 0
Batch Size: 40 transcript lines per call
Context: 12 previously categorized lines of the same transcript
Input: manually reviewed transcripts, not anonymized
Output Format: JSON

Prompt

You are an expert annotator trained in venture capital decision-making and entrepreneurial finance.

Your task is to classify Shark Tank transcript lines into exactly ONE category.

The classification is based on economically meaningful dimensions used by venture capitalists when evaluating startups.

Allowed categories:
["PRODUCT_TECH", "OPERATIONS_EXECUTION", "COMMERCIAL_STRATEGY", "FOUNDER_TEAM", "FINANCIALS_VALUATION", "INVESTOR_VALUE_ADD", "MARKET_COMPETITION", "NEGOTIATION", "FINAL_COMMENTS", "NON_SUBSTANTIVE"]

Known sharks and guest sharks:
["Alex Rodriguez", "Alli Webb", "Anne Wojcicki", "Ashton Kutcher", "Barbara Corcoran", "Bethenny Frankel", "Blake Mycoskie", "Candace Nelson", "Charles Barkley", "Chris Sacca", "Daniel Lubetzky", "Daymond John", "Emma Grede", "Gwyneth Paltrow", "Holly Cooper", "Jamie Siminoff", "Jason Blum", "Jason Richardson", "Katrina Lake", "Kendra Scott", "Kevin Harrington", "Kevin Hart", "Kevin O'Leary", "Lori Greiner", "Maria Sharapova", "Mark Cuban", "Matt Higgins", "Michael Rubin", "Nirav Tolia", "Peter Jones", "Rohan Oza", "Sara Blakely", "Steve Tisch", "Tony Xu", "Troy Carter"]

Category definitions:
PRODUCT_TECH – What the product or service is and how it works.
Includes: features, functionality, design, ingredients, materials, demo, use case, customer problem, problem-solution explanation, benefits, uniqueness, and product-specific IP.
Rule: If a line mainly explains the product or why it is useful, choose this.

OPERATIONS_EXECUTION – How the product is produced and delivered.
Includes: manufacturing, factories, suppliers, sourcing, production process, logistics, shipping, inventory, capacity, lead times, quality control, packaging, warehousing, and operational scalability.

COMMERCIAL_STRATEGY – How the company reaches customers and generates sales.
Includes: distribution channels, retail, wholesale, e-commerce, website, store rollout, partnerships, subscriptions, licensing channels, branding, marketing, advertising, customer acquisition, and go-to-market strategy.
Rule: Focus on how they sell, where they sell, and how they get customers.

FOUNDER_TEAM – Background and characteristics of the founder(s).
Includes: experience, prior jobs, expertise, credentials, personal story, motivation for founding, origin story, hometown, place of origin, identity/background details, and other personal background information about the entrepreneur(s) or founding team.
Rule: Use this for any background information about the entrepreneur(s) or founding team, even if it is brief or introductory, as long as it is about the person rather than the product or business itself.

FINANCIALS_VALUATION – Financial performance and financial structure.
Includes: initial ask amount, equity offer, valuation, revenue, sales, units sold, pricing, retail price, wholesale price, margins, profit/loss, debt, cash needs, funding history, prior investment, ownership structure, projections, money invested, and use of funds.

INVESTOR_VALUE_ADD – Non-monetary contributions investors can provide.
Includes: network access, branding help, retail access, distribution connections, licensing expertise, manufacturing help, strategic guidance, industry contacts, and introductions.
Rule: Use only when the value-add is discussed outside active deal bargaining.

MARKET_COMPETITION – External market environment and opportunity.
Includes: market size, customer segments, competitors, substitutes, demand, trends, barriers to entry, industry dynamics, defensibility at the market level, and scalability of demand.

NEGOTIATION – Active deal-making and bargaining.
Includes: offers, counteroffers, changes in equity, valuation negotiation, royalties, loans, deal

terms, contingencies, and explicit conditions.
Rule: ONLY use this when actual bargaining or deal terms are being discussed. The initial pitch ask is NOT negotiation.

FINAL_COMMENTS – Final outcome and closing remarks after the decision is effectively made.
Includes: "I'm out", final reasoning after declining, acceptance of an offer, immediate reactions after a deal is made or rejected, congratulations, thanks, celebratory closing remarks, and immediate post-decision reflections.
Rule: Once the final outcome is clearly decided, directly related closing lines should usually remain in FINAL_COMMENTS until the scene clearly shifts to a new topic or separate post-scene discussion.

NON_SUBSTANTIVE – Fallback category for lines that do not fit any other category.
Includes: greetings, applause, jokes, laughter, banter, compliments, tasting reactions, surprise reactions, filler talk, and casual remarks with no substantive business content.
Rule: Use NON_SUBSTANTIVE only if no other category is a better fit. Always try to assign the line to one of the substantive categories, or to FINAL_COMMENTS if it belongs to the closing sequence, before using NON_SUBSTANTIVE.

General principles:
1. Assign exactly ONE category from the allowed list to every line.
2. Use the line itself plus nearby context.
3. Choose the MOST SPECIFIC and PRIMARY function of the line.
4. Prefer specific categories over general ones.
5. Do not explain reasoning.
6. Return only valid JSON. No prose, no markdown, no explanations.

Context continuity rules:
7. If an investor asks a question in a category, the entrepreneur's response should usually stay in that same category unless the topic clearly changes.
8. If an entrepreneur answers across multiple short consecutive lines, keep the same category until the topic clearly changes.
9. If a line is only a linguistic bridge, filler transition, or setup line with little standalone meaning, classify it according to the surrounding substantive context.

Critical boundary rules:
10. The entrepreneur's INITIAL ASK, for example "I am seeking $100,000 for 10% equity", always belongs to FINANCIALS_VALUATION, not NEGOTIATION.
11. NEGOTIATION starts only when active bargaining begins: an actual offer, counteroffer, equity adjustment, royalty, loan, contingency, or explicit deal term.
12. Once an active offer or counteroffer sequence begins, all directly related lines remain NEGOTIATION until that bargaining thread clearly ends, even if investor value-add is mentioned as part of the offer.
13. INVESTOR_VALUE_ADD is only for non-monetary investor help outside active bargaining.
14. FOUNDER_TEAM is only about the entrepreneur, founding team, or company origin in a biographical sense.
15. If a KNOWN SHARK talks about their own experience, contacts, background, or expertise, do NOT classify that as FOUNDER_TEAM. Use INVESTOR_VALUE_ADD if it is substantive help, NEGOTIATION if it is part of a live deal offer, or NON_SUBSTANTIVE if it is only casual talk.
16. If a line mainly explains the customer problem, use case, why the solution is needed, or how the product works, prefer PRODUCT_TECH over FOUNDER_TEAM.
17. Questions or statements about revenue, units sold, prices, margins, profit, loss, valuation, equity, money invested, prior funding, ownership, purchase orders as traction, or use of funds belong to FINANCIALS_VALUATION.
18. Questions or statements about channels, retail placement, wholesale, online sales, distribution partners, subscriptions, licensing channels, customer acquisition, marketing, branding, or rollout belong to COMMERCIAL_STRATEGY.
19. FINAL_COMMENTS is for clear final decision statements and the immediate closing exchange after the outcome is decided.
20. If a deal is accepted, the acceptance line and the directly related closing lines that follow, such as congratulations, thanks, celebration, and immediate reflections, should usually be labeled

```
FINAL_COMMENTS.
21.  If all sharks are out, the final out-reasons and the immediate closing reflections should
usually be labeled FINAL_COMMENTS.
22.  NON_SUBSTANTIVE should be used for lightweight talk only when it is not functioning as part of
the final closing sequence.
23.  If a line contains both banter and substantive business content, prefer the substantive
business category over NON_SUBSTANTIVE.

Boundary guide:
- PRODUCT_TECH = what the thing is, what problem it solves, how it works
- OPERATIONS_EXECUTION = how it is made and delivered
- COMMERCIAL_STRATEGY = how it is sold and how customers are reached
- FOUNDER_TEAM = who the founder is and why they started it
- FINANCIALS_VALUATION = money facts, traction, ownership, valuation, initial ask
- INVESTOR_VALUE_ADD = non-cash help outside bargaining
- MARKET_COMPETITION = external market context
- NEGOTIATION = live deal bargaining over terms
- FINAL_COMMENTS = final decision, acceptance/rejection, and immediate closing remarks
- NON_SUBSTANTIVE = non-business exchange outside the final closing sequence

Examples:
- "I'm seeking $50,000 for 10% equity." -> "FINANCIALS_VALUATION"
- "What does it cost you to make one unit?" -> "OPERATIONS_EXECUTION"
- "What stores are you in?" -> "COMMERCIAL_STRATEGY"
- "What do you sell it for?" -> "FINANCIALS_VALUATION"
- "How many units have you sold?" -> "FINANCIALS_VALUATION"
- "I started this company after working 15 years as a nurse." -> "FOUNDER_TEAM"
- "My fingers were hurting, so I designed a better keyboard layout." -> "PRODUCT_TECH" unless
surrounding context is clearly founder biography/origin
- "How big is the market?" -> "MARKET_COMPETITION"
- "I'll give you $100,000 for 20%." -> "NEGOTIATION"
- "I can get you into major retailers." -> "INVESTOR_VALUE_ADD"
- "I'm out because I don't think the market is big enough." -> "FINAL_COMMENTS"
- "Let's do it." -> "FINAL_COMMENTS"
- "Congratulations, guys." -> "FINAL_COMMENTS" if it follows the accepted or rejected final outcome
- "Wow, that tastes great." -> "NON_SUBSTANTIVE"

Output format:
[
  {"id":  1, "category":  "PRODUCT_TECH"},
  {"id":  2, "category":  "NEGOTIATION"}
]
```

C.1.1 Example *Shark Tank* Transcript

```
S01_E09_01
YouTube Link: https://youtu.be/RS_NIrX2u2o?si=DiTUA2Pj25x9_mFG
[FOUNDER_TEAM] Dan Mackey: Hi, my name is Dan Mackey, and I am the creator of Chill Soda.
[FINANCIALS_VALUATION] Dan Mackey: I am seeking a minimum investment of $50,000 for 10% equity.
[PRODUCT_TECH] Dan Mackey: Chill is a truly healthy line of soda.
[FOUNDER_TEAM] Dan Mackey: I have some teachers in my family.
[PRODUCT_TECH] Dan Mackey: When California was getting rid of sodas from school campuses, we saw
what they replaced them with, these diet sodas that are full of chemicals and no one wants to
drink, juices full of sugar, and they're like, is this the best they got?
[PRODUCT_TECH] Dan Mackey: If someone wanted to make a boatload of money, they should invent a
healthy soda.
[NON_SUBSTANTIVE] Dan Mackey: Oh.
[FOUNDER_TEAM] Dan Mackey: Guess what I did?
```

[PRODUCT_TECH] Dan Mackey: Chill is uniquely low-glycemic, it's organic, it has no high-fructose corn syrup, no simple sugars, it's nutritious, it's light in calories, it has antioxidants, vitamins, but the best thing?
[PRODUCT_TECH] Dan Mackey: It tastes like regular soda.
[NON_SUBSTANTIVE] Dan Mackey: It is good.
[NON_SUBSTANTIVE] Dan Mackey: I brought a few samples I'd love for you to taste.
[PRODUCT_TECH] Kevin O'Leary Is there an artificial sweetener in here?
[NON_SUBSTANTIVE] Robert Herjavec: Thanks, Dan.
[PRODUCT_TECH] Dan Mackey: There is no artificial sweetener.
[PRODUCT_TECH] Dan Mackey: Because organic means there's no chemicals, nothing artificial whatsoever.
[NON_SUBSTANTIVE] Kevin O'Leary: Thank you.
[PRODUCT_TECH] Dan Mackey: We use something called agave nectar.
[NON_SUBSTANTIVE] Robert Herjavec: Oh, that was pretty good.
[NON_SUBSTANTIVE] Barbara Corcoran: It's good.
[PRODUCT_TECH] Robert Herjavec: And that's nutritious like honey, but it's low glycemic, and it's actually sweeter than sugar, so we're able to get the same sweetness level for less calories.
[FINANCIALS_VALUATION] Robert Herjavec: How many calories in one can?
[FINANCIALS_VALUATION] Dan Mackey: In one can, we have 85 calories.
[NON_SUBSTANTIVE] Robert Herjavec : That's pretty good.
[PRODUCT_TECH] Dan Mackey: Your typical soda is 140, and a lot of the organic sodas on the market are like 160, 180.
[COMMERCIAL_STRATEGY] Kevin O'Leary: Okay, Dan, here's a question for you.
[COMMERCIAL_STRATEGY] Dan Mackey: Sure.
[COMMERCIAL_STRATEGY] Kevin O'Leary: What is the thing that is the ultimate leverage of success for all players in the beverage market?
[COMMERCIAL_STRATEGY] Kevin O'Leary: I hope you know this answer.
[COMMERCIAL_STRATEGY] Dan Mackey: ...You gotta taste good.
[COMMERCIAL_STRATEGY] Kevin O'Leary: No, no, no.
[COMMERCIAL_STRATEGY] Robert Herjavec and Daymond John: No.
[COMMERCIAL_STRATEGY] Daymond John: Incorrect.
[COMMERCIAL_STRATEGY] Kevin O'Leary: What do you need that you don't have right now and is the reason you're standing in front of me?
[COMMERCIAL_STRATEGY] Dan Mackey: Well...
[COMMERCIAL_STRATEGY] Kevin O'Leary: You gotta get this.
[COMMERCIAL_STRATEGY] Dan Mackey: You're gonna have to answer it for me.
[COMMERCIAL_STRATEGY] Kevin O'Leary: Distribution!
[COMMERCIAL_STRATEGY] Robert Herjavec: Shelf space.
[COMMERCIAL_STRATEGY] Kevin Harrington: Shelf space.
[COMMERCIAL_STRATEGY] Dan Mackey: Okay.
[COMMERCIAL_STRATEGY] Dan Mackey: Sorry, I thought that was a given because...
[COMMERCIAL_STRATEGY] Kevin O'Leary: You don't have it.
[COMMERCIAL_STRATEGY] Dan Mackey: Oh, actually I do.
[COMMERCIAL_STRATEGY] Kevin O'Leary: What do you have?
[COMMERCIAL_STRATEGY] Dan Mackey: We actually already have a distributor on board.
[COMMERCIAL_STRATEGY] Dan Mackey: They're in 40 different territories throughout the United States.
[COMMERCIAL_STRATEGY] Dan Mackey: So we're rocking and rolling already.
[COMMERCIAL_STRATEGY] Robert Herjavec: Dan, is there a store that I can go into today and see your product on a shelf?
[COMMERCIAL_STRATEGY] Dan Mackey: There is not because you're getting in on the ground floor of this.
[FINANCIALS_VALUATION] Daymond John: Do you have any sales?
[FINANCIALS_VALUATION] Dan Mackey: We already have sold 250 units and we have... [Distribution and sales strategies] Daymond John: 250 units, cases or pieces?
[FINANCIALS_VALUATION] Dan Mackey: Cans.
[NON_SUBSTANTIVE] Daymond John: ...Okay.
[NON_SUBSTANTIVE] Robert Herjavec: hehe.
[FINANCIALS_VALUATION] Daymond John: How much does 250 cans of soda at approximately 40 cents equal?

[FINANCIALS_VALUATION] Dan Mackey: Well, $175,000.
[NON_SUBSTANTIVE] Kevin Harrington: Whoa, whoa, whoa, whoa, whoa.
[FINANCIALS_VALUATION] Daymond John: How many have you sold?
[FINANCIALS_VALUATION] Dan Mackey: 250,000 cans.
[NON_SUBSTANTIVE] all sharks: Oh, all right.
[FINANCIALS_VALUATION] Barbara Corcoran: Thousand?
[FINANCIALS_VALUATION] Daymond John: You said 250 cans.
[NON_SUBSTANTIVE] Dan Mackey: I'm sorry.
[NON_SUBSTANTIVE] Dan Mackey: Yeah, that's not very much.
[FINANCIALS_VALUATION] Kevin O'Leary: So how many dollars, how many dollars have you collected in sales from that?
[FINANCIALS_VALUATION] Dan Mackey: So from that, we got $175,000.
[NON_SUBSTANTIVE] Robert Herjavec: That's not bad.
[NON_SUBSTANTIVE] Daymond John: Dan, you just gave yourself mouth to mouth.
[COMMERCIAL_STRATEGY] Robert Herjavec: Dan, the $175,000 in sales that you got, you told me that it's not available on any shelf.
[COMMERCIAL_STRATEGY] Robert Herjavec: So where did you sell the $175,000 to?
[COMMERCIAL_STRATEGY] Dan Mackey: That was sold to the distributor.
[OPERATIONS_EXECUTION] Dan Mackey: And they're actually helping us with the production run.
[INVESTOR_VALUE_ADD] Dan Mackey: They are so behind it.
[INVESTOR_VALUE_ADD] Robert Herjavec: Why are they being so unbelievably... nice to you?
[INVESTOR_VALUE_ADD] Dan Mackey: Because they have a great relationship with my beverage management company.
[INVESTOR_VALUE_ADD] Kevin O'Leary: What is a beverage management company?
[INVESTOR_VALUE_ADD] Kevin O'Leary: Beverage management, is that a consultant?
[INVESTOR_VALUE_ADD] Dan Mackey: They are consultants, yeah, because I don't know that much about the beverage industry.
[FOUNDER_TEAM] Daymond John: Ooh, that's not something you should have just said to me.
[FOUNDER_TEAM] Kevin O'Leary: What are you doing here?
[NON_SUBSTANTIVE] Robert Herjavec: hehehe.
[FOUNDER_TEAM] Dan Mackey: Well, no, I mean, I'm...
[FOUNDER_TEAM] Robert Herjavec: Dan, don't you think you should know the beverage business if you're going to devote your life to it?
[NON_SUBSTANTIVE] Daymond John: At least lie to me.
[FOUNDER_TEAM] Dan Mackey: Well... I'm a marketing guy.
[FOUNDER_TEAM] Dan Mackey: I'm not in the beverage industry.
[FOUNDER_TEAM] Daymond John: You know, you should never have ever told me that you do not know that business.
[FINAL_COMMENTS] Daymond John: I am out.
[FINAL_COMMENTS] Kevin Harrington: You're going to need far more than $50,000 to get over the hump here.
[FINAL_COMMENTS] Kevin Harrington: You're going to need millions of dollars.
[FINAL_COMMENTS] Kevin Harrington: And I just got to tell you, uh, you've got a lot of guts, and good luck to you, but I'm out.
[FINAL_COMMENTS] Robert Herjavec: Dan, let me tell you, I think you're a good guy, but I think you're an incredibly naive guy.
[FINAL_COMMENTS] Robert Herjavec: You've got to be in control of your own destiny, and you're not.
[FINAL_COMMENTS] Robert Herjavec: I'm out.
[FINAL_COMMENTS] Kevin O'Leary: When you woke up two years ago and said, I want to go into a business, why did you pick the most brutally competitive consumer goods market in the world?
[FINAL_COMMENTS] Kevin O'Leary: What would make you do that?
[FINAL_COMMENTS] Dan Mackey: Because of the business model behind it.
[FINAL_COMMENTS] Kevin O'Leary: Part of what I have to do every day is to try and find opportunities where I put money in harm's way and I get a return.
[FINAL_COMMENTS] Dan Mackey: Correct.
[FINAL_COMMENTS] Kevin O'Leary: I don't pick places where there are thousands of dead and rotting corpses.
[FINAL_COMMENTS] Kevin O'Leary: That's what the beverage industry is.
[FINAL_COMMENTS] Kevin O'Leary: It's guys like you getting slaughtered.

[FINAL_COMMENTS] Kevin O'Leary: Look, I don't want to depress you or not give you the motivation you need.
[FINAL_COMMENTS] Barbara Corcoran: Who are you kidding?
[FINAL_COMMENTS] Kevin O'Leary: I think it's fraught with risk.
[FINAL_COMMENTS] Kevin O'Leary: I'm out.
[FINAL_COMMENTS] Barbara Corcoran: Don't worry, Dan, you got a lot more hair than this guy has and a lot more enthusiasm.
[FINAL_COMMENTS] Barbara Corcoran: You're a lot more positive in my book.
[FINAL_COMMENTS] Robert Herjavec Are you cutting him a check?
[FINAL_COMMENTS] Barbara Corcoran: I'm going to tell you what I like about this.
[FINAL_COMMENTS] Barbara Corcoran: I love the product.
[FINAL_COMMENTS] Barbara Corcoran: I don't know what to be afraid of, so maybe I'm as naive as he.
[FINAL_COMMENTS] Kevin O'Leary: I guess you'll find out the hard way.
[FINAL_COMMENTS] Barbara Corcoran: He's done it right from the beginning.
[FINAL_COMMENTS] Barbara Corcoran: He's working from his heart.
[FINAL_COMMENTS] Barbara Corcoran: And you want to know what I think the best trade is?
[FINAL_COMMENTS] Barbara Corcoran: You're not slick.
[FINAL_COMMENTS] Barbara Corcoran: You stand there and you answer honestly.
[FINAL_COMMENTS] Robert Herjavec: He knows nothing.
[FINAL_COMMENTS] Barbara Corcoran: Hey, look at this guy.
[FINAL_COMMENTS] Barbara Corcoran: He's a straight shooter.
[FINAL_COMMENTS] Daymond John: Are you giving him an offer, Barbara?
[FINAL_COMMENTS] Dan Mackey: She is the type of person we're targeting.
[FINAL_COMMENTS] Dan Mackey: I mean, we want people who are smart and see the potential in this product.
[FINAL_COMMENTS] Kevin O'Leary: Barbara, please don't give this guy any money.
[FINAL_COMMENTS] Kevin O'Leary: No!
[FINAL_COMMENTS] Robert Herjavec: Don't do it.
[FINAL_COMMENTS] Barbara Corcoran: I am gonna write you a check.
[FINAL_COMMENTS] Kevin O'Leary: No.
[FINAL_COMMENTS] Daymond John: Are you crazy?
[FINAL_COMMENTS] Barbara Corcoran: He's a thoughtful guy.
[FINAL_COMMENTS] Barbara Corcoran: He's careful with what he's doing.
[FINAL_COMMENTS] Daymond John: I am too.
[FINAL_COMMENTS] Daymond John: Give me the money.
[FINAL_COMMENTS] Kevin O'Leary: I'll give you $20,000 for therapy.
[FINAL_COMMENTS] Barbara Corcoran: hahaha.
[FINAL_COMMENTS] Daymond John: Barbara, you're gonna call your partner and you're gonna say, well, I don't know what we're doing, he's gonna go, me neither.
[FINAL_COMMENTS] Robert Herjavec: Don't do it.
[FINAL_COMMENTS] Barbara Corcoran: Please don't lecture me.
[FINAL_COMMENTS] Kevin O'Leary: Just a moment, listen.
[FINAL_COMMENTS] Kevin O'Leary: I, I would never do this, ever.
[FINAL_COMMENTS] Kevin O'Leary: You know that.
[FINAL_COMMENTS] Kevin O'Leary: But I feel a power in the room right now.
[FINAL_COMMENTS] Kevin O'Leary: A force is guiding me, and it's begging you.
[FINAL_COMMENTS] Kevin O'Leary: Don't do it!
[FINAL_COMMENTS] Kevin O'Leary: You must be out of your mind.
[FINAL_COMMENTS] Dan Mackey: I would love to have her on board, because one, I'd love to have a woman's perspective.
[NEGOTIATION] Barbara Corcoran: The offer is for the $50,000, but I need to get 20% of your business.
[FINAL_COMMENTS] Kevin O'Leary: No!
[FINAL_COMMENTS] Kevin O'Leary: Oh, I can't take it.
[FINAL_COMMENTS] Kevin O'Leary: I can't.
[FINAL_COMMENTS] Robert Herjavec: Took a flyer.
[FINAL_COMMENTS] Kevin Harrington: ...Is he in?
[FINAL_COMMENTS] Dan Mackey: ...Barbara,I am in.
[FINAL_COMMENTS] Dan Mackey: I love that you're in.

```
[FINAL_COMMENTS] Kevin O'Leary: The guy is going to get vaporized by the beverage industry.
[FINAL_COMMENTS] Dan Mackey: You're going to be so surprised.
[FINAL_COMMENTS] Kevin O'Leary: Vaporized.
[FINAL_COMMENTS] Kevin Harrington: This is very remarkable.
[FINAL_COMMENTS] Barbara Corcoran: I can't wait to have them green with envy.
[FINAL_COMMENTS] Dan Mackey: I can't either.
[FINAL_COMMENTS] Kevin O'Leary: You are guilty of murdering money.
[FINAL_COMMENTS] Kevin O'Leary: You're going to go to jail.
[FINAL_COMMENTS] Dan Mackey: They are going to be so sorry they didn't get in on the ground floor,
Barbara.
[FINAL_COMMENTS] Dan Mackey: Thank you so much.
[FINAL_COMMENTS] Dan Mackey: You are such a sweetheart.
[FINAL_COMMENTS] Barbara Corcoran: You are cute.
[FINAL_COMMENTS] Barbara Corcoran: It takes a lot to know.
[FINAL_COMMENTS] Kevin O'Leary: Have you lost your mind?
[FINAL_COMMENTS] Barbara Corcoran: Not at all.
[FINAL_COMMENTS] Barbara Corcoran: I can't wait to show you my profits.
[FINAL_COMMENTS] Dan Mackey: Thank you all.
[FINAL_COMMENTS] Dan Mackey: I greatly appreciate it.
[FINAL_COMMENTS] Dan Mackey: Barbara, we'll be in touch.
[FINAL_COMMENTS] Kevin O'Leary: Barbara, you have offended...
[FINAL_COMMENTS] Kevin O'Leary: You have offended the primal forces of capitalism here.
[FINAL_COMMENTS] Barbara Corcoran: Enough.
[FINAL_COMMENTS] Barbara Corcoran: I'm happy with the deal.
[FINAL_COMMENTS] Robert Herjavec: Good luck, Barbara.
[FINAL_COMMENTS] Kevin O'Leary: I need a drink.
[FINAL_COMMENTS] Dan Mackey: I got Barbara on board with exactly what I wanted to have happen, so I
am totally stoked.
```

C.1.2 Example *Shark Tank* Transcript, after cleaning

The same pitch as Appendix C.1.1, after the cleaning described in Section 2: investor turns dropped, the founder's negotiation and closing lines dropped, speaker labels and names replaced. This is the text every variable is computed from. 347 words, against 1,353 in the raw transcript.

```
S01_E09_01  (cleaned)
Hi, my name is FOUNDER, and I am the creator of Company.
I am seeking a minimum investment of $50,000 for 10% equity.
Chill is a truly healthy line of soda.
I have some teachers in my family.
When California was getting rid of sodas from school campuses, we saw what they replaced them with,
these diet sodas that are full of chemicals and no one wants to drink, juices full of sugar, and
they're like, is this the best they got?
If someone wanted to make a boatload of money, they should invent a healthy soda.
Oh.
Guess what I did?
Chill is uniquely low-glycemic, it's organic, it has no high-fructose corn syrup, no simple sugars,
it's nutritious, it's light in calories, it has antioxidants, vitamins, but the best thing?
It tastes like regular soda.
It is good.
I brought a few samples I'd love for you to taste.
There is no artificial sweetener.
Because organic means there's no chemicals, nothing artificial whatsoever.
We use something called agave nectar.
In one can, we have 85 calories.
Your typical soda is 140, and a lot of the organic sodas on the market are like 160, 180.
Sure.
...You gotta taste good.
```

```
Well...
You're gonna have to answer it for me.
Okay.
Sorry, I thought that was a given because...
Oh, actually I do.
We actually already have a distributor on board.
They're in 40 different territories throughout the United States.
So we're rocking and rolling already.
There is not because you're getting in on the ground floor of this.
We already have sold 250 units and we have...
Cans.
Well, $175,000.
250,000 cans.
I'm sorry.
Yeah, that's not very much.
So from that, we got $175,000.
That was sold to the distributor.
And they're actually helping us with the production run.
They are so behind it.
Because they have a great relationship with my beverage management company.
They are consultants, yeah, because I don't know that much about the beverage industry.
Well, no, I mean, I'm...
Well... I'm a marketing guy.
I'm not in the beverage industry.
```

C.2 Measurement

C.2.1 Hard finance (H_{finance}) and soft finance (S_{finance})

The 28 finance variables are defined in advance with the codebook below. Claude extracted each variable from the cleaned transcript using this codebook, and every extracted value was checked by hand against the transcript. Both indices come from the same extraction and the same standard of recording; they differ in what the figure refers to. Hard finance holds the 21 realized figures, each describing something the venture has already done. Soft finance holds the 7 forward claims: the projections and the ask, whose content is a claim about what has not happened. Table C.14 gives the sign each figure carries in its index.

`ask_amount` *(Numeric; Ask)* Total amount of capital requested by the entrepreneur from investors during the pitch, expressed in USD. Extracted by Claude using the codebook; checked by hand against the transcript.

`bootstrapped` *(Binary; Realized)* Indicates whether the startup has been funded WITHOUT outside equity investors up to the moment of the pitch (1 = Yes / Bootstrapped, 0 = No / Not bootstrapped). Keep empty only if funding history is not discussed at all. Decision rule — code 1 (Yes) if ALL of the following are true: (a) No equity has been sold to angels, VCs, accelerators, strategic investors, or any institutional/professional investor; (b) The company has been financed by some combination of: founder savings, founder credit cards, founder personal loans, revenue / reinvested profits, bank loans, SBA loans, or loans/gifts from friends and family that did NOT involve giving up equity. Code 0 (No) if ANY equity has previously been sold to an outside investor (angel, VC, accelerator, strategic, crowdfunding-for-equity), or if a convertible note / SAFE has been issued. Debt alone does NOT disqualify bootstrapping. Important cues in Shark Tank dialogue: - 'I've self-funded this', 'put in my own money', 'no outside investors', 'maxed out my credit cards', 'took out a second mortgage' → 1 - 'we raised a seed round', 'angel invested $X for Y%', 'we're backed by [VC]', 'previous investor owns X%' → 0 - Friends-and-family money: 1 if it was a loan/gift, 0 if they got equity for it. - If only the current Shark Tank ask is on the table and no prior equity investor exists, code 1. Extracted by Claude using the codebook; checked by hand against the transcript.

`customer_acquisition_cost` *(Numeric; Realized)* Average cost to acquire one paying customer. Calculated as total sales and marketing expenses divided by the number of new customers acquired in the same period. Extracted by Claude using the codebook; checked by hand against the transcript.

`debt_amount` *(Numeric; Realized)* Total outstanding debt obligations at the time of the pitch, including loans, convertible notes, credit lines, and other liabilities. Expressed in USD. Extracted by Claude using the codebook; checked

Table C.14: The sign each finance figure would carry by hand, against the sign the data gives

Variable	What it records	Hand	Data	Stable
Panel A. Hard finance, the realized figures				
gross_margin	gross margin, stated or computed	+1	+1	100%
is_profitable	profitable in the latest period	+1	+1	100%
revenue_current	revenue to date	+1	+1	100%
units_sold	units sold to date	+1	+1	100%
founder_capital_invested	founders' own money in the business	+1	−1	100%
revenue_growth_past	past revenue growth	+1	−1	95%
gross_margin_stated	gross margin as stated	+1	+1	94%
development_expense	spent on product development	+1	+1	62%
unit_price	price per unit	—	−1	100%
marketing_spend	spent on marketing	—	−1	100%
funding_raised	external funding raised to date	—	−1	96%
has_marketing_spend	spends on marketing at all	—	−1	92%
bootstrapped	no outside equity taken	—	+1	85%
prior_investment	had a professional investor	—	−1	71%
nonfounder_equity_share	equity held by non-founders	—	−1	66%
debt_amount	debt outstanding	−1	−1	100%
has_debt	carries debt	−1	−1	100%
unit_cost	cost to make one unit	−1	−1	99%
fixed_costs_monthly	monthly fixed costs	−1	+1	99%
customer_acquisition_cost	cost of acquiring one customer	−1	−1	98%
personal_liability	founders personally liable	−1	−1	67%
Panel B. Soft finance, the forward claims				
ask_amount	capital requested	+1	−1	100%
market_size_customers	market size in customers	+1	+1	100%
implied_valuation	valuation the ask implies	+1	−1	100%
revenue_expected	revenue projected	+1	−1	87%
market_size_revenue	market size in revenue	+1	+1	67%
payback_period	months to recover the cost	−1	−1	99%
equity_ask	equity offered in exchange	−1	−1	95%

Notes. The sign multiplies the standardized figure before the figures are averaged, so a +1 means a larger value raises the index and a −1 means it lowers it. *Hand* is the sign the meaning of the figure would give it, with — marking the seven whose direction is contestable enough that we would not assign one. *Data* is the sign the construction of Section 4.5 gives it, and *Stable* is the share of the 100 fold estimates that agree with it. The indices used in the paper are built on the *Data* column.

Of the 21 hand-signable stated figures the data picks the opposite sign for six, and five of those six are stable at 95 percent of fold estimates or better: the founders' own capital, the monthly fixed costs, past revenue growth, the ask, the implied valuation and the revenue projected. The two constructions are therefore not the same variable. Built each way, the hard finance indices correlate at 0.54 and the soft finance indices at −0.07. Section 6.9 reports what the paper's tables look like under the hand-signed alternative.

Each figure also carries a not-stated indicator wherever more than 30 percent of training pitches leave it blank, 15 of them in hard finance and 4 in soft finance, and these are signed by the same rule as the figures themselves. Hard finance therefore averages all 36 of its columns, 21 figures and 15 indicators, and soft finance all 11, 7 figures and 4 indicators.

Seven of the hard finance figures are marked — in the *Hand* column because they can describe either a stronger or a weaker venture depending on what else is true. Prior outside investment is validation and it is also dilution; marketing spend is demand being built and it is also a venture that cannot sell without it. They enter the index like every other figure, since the construction asserts no direction, and each is also reported separately in Table 7.

by hand against the transcript.

`development_expense` *(Numeric; Realized)* Total product or technology development cost in USD Extracted by Claude using the codebook; checked by hand against the transcript.

`equity_ask` *(Numeric; Ask)* Percentage of company equity the founder is offering in exchange for the requested investment. Extracted by Claude using the codebook; checked by hand against the transcript.

`fixed_costs_monthly` *(Numeric; Realized)* Total monthly fixed operating costs / overhead in USD. Captures the recurring costs the business pays each month regardless of sales volume — sometimes called 'overhead', 'burn', 'monthly nut', or 'operating expenses'. Include: - Rent, utilities, insurance - Salaries and payroll for employees and founders (when stated) - Software subscriptions, recurring services - Loan payments / debt servicing (recurring portion) - Any cost the founder describes as 'overhead', 'fixed costs', 'monthly expenses', or 'what it costs us to keep the lights on' Do NOT include: - Cost of goods sold / variable per-unit costs (those are unit_cost) - Marketing spend (that's marketing_spend — kept separate even though it can be recurring) - One-time costs like product development, equipment purchase, legal fees for incorporation (those go in development_expense or are not captured) Period normalization — always express as MONTHLY: - 'Our overhead is $21K a month' → 21,000 - 'We spend $250K a year on overhead' → 250,000 ÷ 12 ≈ 20,833 - 'Weekly burn is $5K' → 5,000 × 4.33 ≈ 21,650 - 'Quarterly OpEx of $90K' → 90,000 ÷ 3 = 30,000 Common Shark Tank cues: - 'Our monthly overhead is $X', 'we burn $X a month', 'fixed costs are $X', 'rent and salaries come to $X', 'it costs us $X a month to run the business'. - If the founder lists individual line items ('rent is $5K, payroll is $15K, insurance is $1K'), sum them. Keep empty if the founder never discusses overhead or operating costs in any form. Extracted by Claude using the codebook; checked by hand against the transcript.

`founder_capital_invested` *(Numeric; Realized)* Total amount of the founders' OWN money invested into the business to date, expressed in USD. Captures founder financial 'skin in the game' — money the founders personally put into the company from their own savings, personal credit, or personal borrowing. Include: - Founder personal savings put into the business - Personal credit card debt or personal loans the founder took out and channeled into the company - Second mortgages, retirement account withdrawals (401k cash-outs), or any other personal funds redirected to fund the business - Money from co-founders' own pockets (sum across all founders) Do NOT include: - Money raised from outside investors (angels, VCs, accelerators) — that goes in funding_raised - Bank loans or SBA loans taken in the company's name (those go in debt_amount) - Friends-and-family money UNLESS the founder explicitly counts it as their own contribution - Revenue reinvested into the business (that's bootstrapping via revenue, not founder capital) - Unpaid founder salary / sweat equity (this variable is dollars actually put in, not opportunity cost) Common Shark Tank cues: - 'I invested $400,000 of my own money', 'put in $50K of my savings', 'we've personally invested $X', 'between the two of us we've put in $Y' → use the stated amount - 'I maxed out my credit cards for $80K to start this' → 80,000 (this also implies has_debt = 1 and debt_amount of at least $80K) - 'I took out a second mortgage for $200K' → 200,000 Keep empty if the founder never quantifies their personal investment, even if they imply self-funding (e.g., 'I bootstrapped this' alone is not enough — need a dollar figure). Extracted by Claude using the codebook; checked by hand against the transcript.

`funding_raised` *(Numeric; Realized)* Total cumulative external funding secured by the startup across all funding rounds to date, expressed in USD. Extracted by Claude using the codebook; checked by hand against the transcript.

`gross_margin` *(Numeric; Realized)* Percentage of revenue remaining after deducting cost of goods sold (COGS). Range: 0–100 (a percentage, not a decimal). Priority order for extraction: 1. If the founder states it directly ('our margins are 60%', 'we make 70 cents on the dollar', 'gross margin is 55%') → use that number. 2. If unit_price AND unit_cost are available (even when GM is not explicitly stated), ALWAYS compute it: gross_margin = ((unit_price - unit_cost) ÷ unit_price) × 100 Example: sells for $20, costs $5 to make → gross_margin = 75. 3. If revenue and COGS (or 'cost of goods') are stated at company level, compute: ((Revenue - COGS) ÷ Revenue) × 100. Notes for Shark Tank dialogue: - 'Landed cost' or 'cost to make' or 'manufacturing cost' = unit_cost for this calculation. - 'Wholesale price' vs 'retail price': if the company sells wholesale, use wholesale price as unit_price. If they sell direct-to-consumer, use retail. - If both wholesale and retail margins are mentioned, prefer the channel that represents the majority of their sales; if unclear, prefer retail/DTC. - Do NOT confuse gross margin with markup. Markup = (price - cost) ÷ cost; gross margin = (price - cost) ÷ price. Always report gross margin. Keep empty ONLY if neither an explicit margin nor unit_price+unit_cost nor revenue+COGS is available. Extracted by Claude using the codebook; checked by hand against the transcript.

`gross_margin_stated` *(Numeric; Realized)* Gross margin as stated by the founder, in percent of revenue (e.g. 'our margins are 60%' → 60). Keep empty if the founder does not state a margin; do not compute it from prices and costs (that is gross_margin). Extracted by Claude using the codebook; checked by hand against the transcript.

`has_debt` *(Binary; Realized)* Indicates whether the company or its founders currently carry any business-related debt

at the time of the pitch (1 = Yes, 0 = No). Keep empty ONLY if debt is never discussed in any form. Code 1 (Yes) if any of the following are mentioned as currently outstanding: - Bank loans, SBA loans, lines of credit, business credit cards with balances - Personal loans or personal credit card debt that the founder took on TO FUND the business (e.g., 'I maxed out my credit cards to launch this', 'second mortgage to start the company') - Loans from friends or family (where repayment is expected — not gifts and not in exchange for equity) - Convertible notes or SAFEs that have not yet converted to equity - Equipment financing, inventory financing, factoring, merchant cash advances - Any phrase like 'we owe $X', 'we're carrying $X in debt', 'still paying off...' Code 0 (No) if: - The founder explicitly says they have no debt ('we have zero debt', 'no loans', 'debt-free') - All prior debt has been fully repaid by the time of the pitch - The only outside capital was equity investment (angel/VC) with no debt component Important distinctions: - Personal debt UNRELATED to the business (e.g., founder's mortgage on their home, student loans, car loan) does NOT count — only debt tied to funding or operating the company. If unclear whether personal debt was used for the business, lean toward 0 unless the founder makes the connection explicit. - Accounts payable / normal trade payables in the regular course of business do NOT count as debt for this variable. - Equity given to investors is NOT debt → does not count here. - If the founder is ASKING for a loan from the Sharks (rather than equity), that is the current ask, not existing debt → does not count toward has_debt unless they also mention pre-existing debt. Consistency check: has_debt must agree with debt_amount. If debt_amount > 0, has_debt = 1. If has_debt = 1, debt_amount should be populated when a figure is given (otherwise leave debt_amount empty but keep has_debt = 1). Extracted by Claude using the codebook; checked by hand against the transcript.

`has_marketing_spend` *(Binary; Realized)* If company spend in marketing (1 = Yes, 0 = No). Keep empty if it is missing Extracted by Claude using the codebook; checked by hand against the transcript.

`implied_valuation` *(Numeric; Ask)* Post-money valuation implied by the pitch terms, computed as ask amount divided by (equity ask / 100), in USD. Computed from the extracted ask and equity offered.

`is_profitable` *(Binary; Realized)* Indicates whether the company is profitable BASED ON THE MOST RECENT PERIOD the founder discusses (1 = Yes, 0 = No). Keep empty if profitability is never addressed. Always anchor the answer to the most recent year / quarter / trailing-12-months the founder mentions. Ignore older statuses if a newer one is given. Examples: - 'We were profitable in 2022 but lost money in 2023' → 0 (use 2023, the most recent) - 'We've been losing money but turned profitable last quarter' → 1 (use last quarter) - 'We made $200K profit last year and are on track for the same this year' → 1 - 'Profitable for three years running' → 1 - 'We're pre-revenue' or 'haven't launched yet' → 0 Profitable means net income > 0 (revenue exceeds ALL costs, not just COGS). Positive gross margin alone is NOT profitability. If the founder only mentions gross profit / gross margin without addressing overall net profit or losses, leave empty rather than infer. Treat 'breaking even' as 0. Treat 'cash flow positive' as 1 only if the founder uses it to mean overall profitability; if it's clearly just operating cash flow while the business is still losing money on paper, leave empty. Extracted by Claude using the codebook; checked by hand against the transcript.

`market_size_customers` *(Numeric; Projected)* Total addressable market in number of customers Extracted by Claude using the codebook; checked by hand against the transcript.

`market_size_revenue` *(Numeric; Projected)* Total addressable market in USD Extracted by Claude using the codebook; checked by hand against the transcript.

`marketing_spend` *(Numeric; Realized)* Total marketing spend in USD Extracted by Claude using the codebook; checked by hand against the transcript.

`nonfounder_equity_share` *(Numeric; Realized)* Percentage of equity held by non-founders (0–100). Anything from previous fundings, or advisor, mentors Extracted by Claude using the codebook; checked by hand against the transcript.

`payback_period` *(Numeric; Projected)* Time (in MONTHS) required to recover the cost of acquiring a customer or the cost of producing/launching the product, from the revenue/margin that customer or product generates. Express as months (convert years × 12). Priority order for extraction: 1. If the founder states it directly ('we pay back in 4 months', 'CAC payback is one year') → use that, converted to months. 2. If not stated but customer_acquisition_cost AND a per-customer contribution are available, compute: payback_period = customer_acquisition_cost ÷ (monthly_revenue_per_custome × gross_margin_decimal) 3. If only unit economics are given (unit_price, unit_cost, units_sold timing), compute approximately: per-unit gross profit = unit_price - unit_cost payback_period (months) ≈ marketing_spend ÷ (per-unit gross profit × units_sold_per_month) 4. If only expected annual revenue and gross margin are given alongside the ask amount or development expense, compute as a rough investment-payback proxy: payback_period (months) ≈ (ask_amount or development_expense) ÷ ((revenue_expected × gross_margin_decimal) ÷ 12) Always show the computation reasoning briefly if you derived it. Keep empty ONLY if none of the above inputs are available. Never leave blank just because the founder didn't use the phrase 'payback period'. Extracted by Claude using the codebook;

checked by hand against the transcript. Because rule 4 uses expected revenue, the variable is grouped with the projected figures.

`personal_liability` *(Binary; Realized)* If founders have personal financial liability like bank loan, credit card loan (1 = Yes, 0 = No). Keep empty if it is missing Extracted by Claude using the codebook; checked by hand against the transcript.

`prior_investment` *(Binary; Realized)* Indicates whether the company has previously received funding from angel investor, VC investor, institutional, or professional investors (1 = Yes, 0 = No). Keep empty if they do not mention it Extracted by Claude using the codebook; checked by hand against the transcript.

`revenue_current` *(Numeric; Realized)* Current or most recent revenue figure in USD Extracted by Claude using the codebook; checked by hand against the transcript.

`revenue_expected` *(Numeric; Projected)* Projected future revenue in USD Extracted by Claude using the codebook; checked by hand against the transcript.

`revenue_growth_past` *(Numeric; Realized)* Historical revenue growth as a percentage (0–100+, can exceed 100 for fast growth). Represents how much revenue grew from one period to the next. Priority order for extraction: 1. If the founder states it directly ('we grew 300% last year', 'revenue tripled', 'doubled year over year') → use that figure. Note: 'tripled' = 200% growth, 'doubled' = 100% growth, '3x' = 200% growth. 2. If the founder gives revenue figures for two or more periods, ALWAYS compute it: growth % = ((latest_revenue - prior_revenue) ÷ prior_revenue) × 100 Example: '$170K, then $540K, then $5M' → use the most recent transition: ((5,000,000 - 540,000) ÷ 540,000) × 100 ≈ 826%. 3. If three or more periods are given and growth is uneven, prefer the most recent period-over-period growth rate (latest vs. immediately prior). Optionally note the multi-year CAGR in the reasoning, but the recorded value is the latest year-over-year rate. Notes for Shark Tank dialogue: - Founders often list revenue by year ('Year 1 we did $X, Year 2 we did $Y, this year we'll do $Z') — treat this as a sequence and compute the latest transition. - Distinguish historical growth (already happened) from projected growth ('we expect to grow 5x next year'). Projected growth belongs in revenue_expected, NOT here. - If only one revenue figure is ever given, leave empty. Do not infer growth from a single data point. Always show the computation briefly when derived. Never leave blank if two or more historical revenue figures are present in the dialogue. Extracted by Claude using the codebook; checked by hand against the transcript.

`unit_cost` *(Numeric; Realized)* Cost to produce one unit in USD Extracted by Claude using the codebook; checked by hand against the transcript.

`unit_price` *(Numeric; Realized)* Price per unit sold or product in USD Extracted by Claude using the codebook; checked by hand against the transcript.

`units_sold` *(Numeric; Realized)* Number of units sold Extracted by Claude using the codebook; checked by hand against the transcript.

Missingness indicators, generated for variables with more than 30 percent missing values in the training sample. The list is for the fixed split and can differ slightly across splits.

`customer_acquisition_cost__missing` *(Binary)* Indicator equal to one if customer acquisition cost is not stated in the pitch.

`debt_amount__missing` *(Binary)* Indicator equal to one if debt amount is not stated in the pitch.

`development_expense__missing` *(Binary)* Indicator equal to one if development expense is not stated in the pitch.

`fixed_costs_monthly__missing` *(Binary)* Indicator equal to one if fixed costs monthly is not stated in the pitch.

`founder_capital_invested__missing` *(Binary)* Indicator equal to one if founder capital invested is not stated in the pitch.

`funding_raised__missing` *(Binary)* Indicator equal to one if funding raised is not stated in the pitch.

`gross_margin_stated__missing` *(Binary)* Indicator equal to one if gross margin stated is not stated in the pitch.

`has_debt__missing` *(Binary)* Indicator equal to one if has debt is not stated in the pitch.

`has_marketing_spend__missing` *(Binary)* Indicator equal to one if has marketing spend is not stated in the pitch.

`is_profitable__missing` *(Binary)* Indicator equal to one if is profitable is not stated in the pitch.

`market_size_customers__missing` *(Binary)* Indicator equal to one if market size customers is not stated in the pitch.

`market_size_revenue__missing` *(Binary)* Indicator equal to one if market size revenue is not stated in the pitch.

`marketing_spend__missing` *(Binary)* Indicator equal to one if marketing spend is not stated in the pitch.

`nonfounder_equity_share__missing` *(Binary)* Indicator equal to one if nonfounder equity share is not stated in the pitch.

`payback_period__missing` *(Binary)* Indicator equal to one if payback period is not stated in the pitch.

`personal_liability__missing` *(Binary)* Indicator equal to one if personal liability is not stated in the pitch.

`revenue_expected__missing` *(Binary)* Indicator equal to one if revenue expected is not stated in the pitch.

`revenue_growth_past__missing` *(Binary)* Indicator equal to one if revenue growth past is not stated in the pitch.

`units_sold__missing` *(Binary)* Indicator equal to one if units sold is not stated in the pitch.

C.2.2 Language variables (S_{language})

The 192 variables of the language set, computed from the founder's statements in the cleaned transcript (negotiation and closing remarks removed) by fixed parsers and dictionaries; no language model is involved. Variables are listed by family, the unit at which the index is built. A rate per thousand words counts dictionary matches and divides by transcript length; a mean averages a norm over the matching words; a parser statistic is computed from the text itself.

Tone and emotion (73)

`Affect` LIWC-22 (Boyd et al., 2022); count or score.
`Negative` LIWC-22 (Boyd et al., 2022); count or score.
`Optimistic` LIWC-22 (Boyd et al., 2022); count or score.
`Positive` LIWC-22 (Boyd et al., 2022); count or score.
`Tone` LIWC-22 (Boyd et al., 2022); count or score.
`conflict` LIWC-22 (Boyd et al., 2022); count or score.
`emo_anger` LIWC-22 (Boyd et al., 2022); count or score.
`emo_anger_per1k` NRC emotion lexicon; rate per thousand words.
`emo_anticipation_per1k` NRC emotion lexicon; rate per thousand words.
`emo_anx` LIWC-22 (Boyd et al., 2022); count or score.
`emo_disgust_per1k` NRC emotion lexicon; rate per thousand words.
`emo_fear_per1k` NRC emotion lexicon; rate per thousand words.
`emo_joy_per1k` NRC emotion lexicon; rate per thousand words.
`emo_neg` LIWC-22 (Boyd et al., 2022); count or score.
`emo_negative_per1k` NRC emotion lexicon; rate per thousand words.
`emo_pos` LIWC-22 (Boyd et al., 2022); count or score.
`emo_positive_per1k` NRC emotion lexicon; rate per thousand words.
`emo_sad` LIWC-22 (Boyd et al., 2022); count or score.
`emo_sadness_per1k` NRC emotion lexicon; rate per thousand words.
`emo_surprise_per1k` NRC emotion lexicon; rate per thousand words.
`emo_trust_per1k` NRC emotion lexicon; rate per thousand words.
`emoint_anger_mean` NRC emotion-intensity lexicon; mean over the words that match.
`emoint_anger_per1k` NRC emotion-intensity lexicon; rate per thousand words.
`emoint_anticipation_mean` NRC emotion-intensity lexicon; mean over the words that match.
`emoint_anticipation_per1k` NRC emotion-intensity lexicon; rate per thousand words.
`emoint_disgust_mean` NRC emotion-intensity lexicon; mean over the words that match.
`emoint_disgust_per1k` NRC emotion-intensity lexicon; rate per thousand words.
`emoint_fear_mean` NRC emotion-intensity lexicon; mean over the words that match.
`emoint_fear_per1k` NRC emotion-intensity lexicon; rate per thousand words.
`emoint_joy_mean` NRC emotion-intensity lexicon; mean over the words that match.
`emoint_joy_per1k` NRC emotion-intensity lexicon; rate per thousand words.

`emoint_sadness_mean` NRC emotion-intensity lexicon; mean over the words that match.
`emoint_sadness_per1k` NRC emotion-intensity lexicon; rate per thousand words.
`emoint_surprise_mean` NRC emotion-intensity lexicon; mean over the words that match.
`emoint_surprise_per1k` NRC emotion-intensity lexicon; rate per thousand words.
`emoint_trust_mean` NRC emotion-intensity lexicon; mean over the words that match.
`emoint_trust_per1k` NRC emotion-intensity lexicon; rate per thousand words.
`emotion` LIWC-22 (Boyd et al., 2022); count or score.
`empath_anger_z` Empath (Fast et al., 2016); standardized score.
`empath_fear_z` Empath (Fast et al., 2016); standardized score.
`empath_negative_emotion_z` Empath (Fast et al., 2016); standardized score.
`empath_positive_emotion_z` Empath (Fast et al., 2016); standardized score.
`empath_sadness_z` Empath (Fast et al., 2016); standardized score.
`fatigue` LIWC-22 (Boyd et al., 2022); count or score.
`feeling` LIWC-22 (Boyd et al., 2022); count or score.
`first_sentence_sentiment` VADER (Hutto and Gilbert, 2014); count or score.
`lack` LIWC-22 (Boyd et al., 2022); count or score.
`last_sentence_sentiment` VADER (Hutto and Gilbert, 2014); count or score.
`lm_negative_neg_per1k` Loughran–McDonald financial dictionary (Loughran and McDonald, 2011); rate per thousand words.
`lm_negative_pos_per1k` Loughran–McDonald financial dictionary (Loughran and McDonald, 2011); rate per thousand words.
`lm_positive_neg_per1k` Loughran–McDonald financial dictionary (Loughran and McDonald, 2011); rate per thousand words.
`lm_positive_pos_per1k` Loughran–McDonald financial dictionary (Loughran and McDonald, 2011); rate per thousand words.
`net_tone` Composite of other variables in this set; count or score.
`nrc_arousal_high_per1k` NRC valence-arousal-dominance norms; rate per thousand words.
`nrc_arousal_mean` NRC valence-arousal-dominance norms; mean over the words that match.
`nrc_dominance_mean` NRC valence-arousal-dominance norms; mean over the words that match.
`nrc_vad_coverage` NRC valence-arousal-dominance norms; count or score.
`nrc_valence_mean` NRC valence-arousal-dominance norms; mean over the words that match.
`pct_negative_sentences` VADER (Hutto and Gilbert, 2014); count or score.
`pct_positive_sentences` VADER (Hutto and Gilbert, 2014); count or score.
`sentential_volatility` Recovered from the project's earlier feature set; count or score.
`sentiment_mean` VADER (Hutto and Gilbert, 2014); mean over the words that match.
`sentiment_sd` VADER (Hutto and Gilbert, 2014); count or score.
`swear` LIWC-22 (Boyd et al., 2022); count or score.
`tone_neg` LIWC-22 (Boyd et al., 2022); count or score.
`vader_compound` VADER (Hutto and Gilbert, 2014); count or score.
`vader_compound_z` VADER (Hutto and Gilbert, 2014); standardized score.
`vader_neg` VADER (Hutto and Gilbert, 2014); count or score.
`vader_neg_z` VADER (Hutto and Gilbert, 2014); standardized score.
`vader_neu` VADER (Hutto and Gilbert, 2014); count or score.
`vader_neu_z` VADER (Hutto and Gilbert, 2014); standardized score.
`vader_pos` VADER (Hutto and Gilbert, 2014); count or score.
`vader_pos_z` VADER (Hutto and Gilbert, 2014); standardized score.

Certainty and hedging (21)

`Constraining` LIWC-22 (Boyd et al., 2022); count or score.
`Litigious` LIWC-22 (Boyd et al., 2022); count or score.
`Risk_minisatation_transference` LIWC-22 (Boyd et al., 2022); count or score.

`StrongModal` LIWC-22 (Boyd et al., 2022); count or score.

`Uncertainty` LIWC-22 (Boyd et al., 2022); count or score.

`WeakModal` LIWC-22 (Boyd et al., 2022); count or score.

`certitude` LIWC-22 (Boyd et al., 2022); count or score.

`discrep` LIWC-22 (Boyd et al., 2022); count or score.

`lm_Uncertainty` LIWC-22 (Boyd et al., 2022); count or score.

`lm_constraining_neg_per1k` Loughran–McDonald financial dictionary (Loughran and McDonald, 2011); rate per thousand words.

`lm_constraining_pos_per1k` Loughran–McDonald financial dictionary (Loughran and McDonald, 2011); rate per thousand words.

`lm_litigious_neg_per1k` Loughran–McDonald financial dictionary (Loughran and McDonald, 2011); rate per thousand words.

`lm_litigious_pos_per1k` Loughran–McDonald financial dictionary (Loughran and McDonald, 2011); rate per thousand words.

`lm_strong_modal_neg_per1k` Loughran–McDonald financial dictionary (Loughran and McDonald, 2011); rate per thousand words.

`lm_strong_modal_pos_per1k` Loughran–McDonald financial dictionary (Loughran and McDonald, 2011); rate per thousand words.

`lm_uncertainty_neg_per1k` Loughran–McDonald financial dictionary (Loughran and McDonald, 2011); rate per thousand words.

`lm_uncertainty_pos_per1k` Loughran–McDonald financial dictionary (Loughran and McDonald, 2011); rate per thousand words.

`lm_weak_modal_neg_per1k` Loughran–McDonald financial dictionary (Loughran and McDonald, 2011); rate per thousand words.

`lm_weak_modal_pos_per1k` Loughran–McDonald financial dictionary (Loughran and McDonald, 2011); rate per thousand words.

`negate` LIWC-22 (Boyd et al., 2022); count or score.

`tentat` LIWC-22 (Boyd et al., 2022); count or score.

Concreteness and numeracy (10)

`dollar_rate` Recovered from the project's earlier feature set; ratio.

`dollars` Parser statistic of the text itself; no dictionary; count or score.

`millions` Parser statistic of the text itself; no dictionary; count or score.

`money` LIWC-22 (Boyd et al., 2022); count or score.

`number` LIWC-22 (Boyd et al., 2022); count or score.

`numbers` Parser statistic of the text itself; no dictionary; count or score.

`percents` Parser statistic of the text itself; no dictionary; count or score.

`quantity` LIWC-22 (Boyd et al., 2022); count or score.

`space` LIWC-22 (Boyd et al., 2022); count or score.

`time` LIWC-22 (Boyd et al., 2022); count or score.

Warmth and competence (27)

`Clout` LIWC-22 (Boyd et al., 2022); count or score.

`Social` LIWC-22 (Boyd et al., 2022); count or score.

`allure` LIWC-22 (Boyd et al., 2022); count or score.

`attention` LIWC-22 (Boyd et al., 2022); count or score.

`bsri_feminine_neg_per1k` Bem Sex-Role Inventory word stems (Bem, 1974); rate per thousand words.

`bsri_feminine_pos_per1k` Bem Sex-Role Inventory word stems (Bem, 1974); rate per thousand words.

`bsri_femininity_pct` Composite of other variables in this set; share.

`bsri_masculine_neg_per1k` Bem Sex-Role Inventory word stems (Bem, 1974); rate per thousand words.

`bsri_masculine_pos_per1k` Bem Sex-Role Inventory word stems (Bem, 1974); rate per thousand words.

`bsri_masculinity_pct` Composite of other variables in this set; share.

`bsri_neutral_neg_per1k` Bem Sex-Role Inventory word stems (Bem, 1974); rate per thousand words.
`bsri_neutral_pos_per1k` Bem Sex-Role Inventory word stems (Bem, 1974); rate per thousand words.
`comm` LIWC-22 (Boyd et al., 2022); count or score.
`eberhardt_ability_neg_per1k` Ability and grindstone word stems; rate per thousand words.
`eberhardt_ability_pos_per1k` Ability and grindstone word stems; rate per thousand words.
`eberhardt_grindstone_neg_per1k` Ability and grindstone word stems; rate per thousand words.
`eberhardt_grindstone_pos_per1k` Ability and grindstone word stems; rate per thousand words.
`moral` LIWC-22 (Boyd et al., 2022); count or score.
`polite` LIWC-22 (Boyd et al., 2022); count or score.
`prosocial` LIWC-22 (Boyd et al., 2022); count or score.
`sad_ability_directed_per1k` SADCAT warmth and competence dictionary (Nicolas et al., 2021); rate per thousand words.
`sad_ability_neg_per1k` SADCAT warmth and competence dictionary (Nicolas et al., 2021); rate per thousand words.
`sad_ability_pos_per1k` SADCAT warmth and competence dictionary (Nicolas et al., 2021); rate per thousand words.
`sad_warmth_directed_per1k` SADCAT warmth and competence dictionary (Nicolas et al., 2021); rate per thousand words.
`sad_warmth_neg_per1k` SADCAT warmth and competence dictionary (Nicolas et al., 2021); rate per thousand words.
`sad_warmth_pos_per1k` SADCAT warmth and competence dictionary (Nicolas et al., 2021); rate per thousand words.
`socbehav` LIWC-22 (Boyd et al., 2022); count or score.

Entrepreneurial orientation (19)

`Creative` LIWC-22 (Boyd et al., 2022); count or score.
`Drives` LIWC-22 (Boyd et al., 2022); count or score.
`EntrepreneurialOrientation` LIWC-22 (Boyd et al., 2022); count or score.
`Guerrilla_skills` LIWC-22 (Boyd et al., 2022); count or score.
`Innovative` LIWC-22 (Boyd et al., 2022); count or score.
`Mentor` LIWC-22 (Boyd et al., 2022); count or score.
`Opportunity` LIWC-22 (Boyd et al., 2022); count or score.
`Perservance_Tenacity` LIWC-22 (Boyd et al., 2022); count or score.
`Proactive` LIWC-22 (Boyd et al., 2022); count or score.
`Reward` LIWC-22 (Boyd et al., 2022); count or score.
`Risk` LIWC-22 (Boyd et al., 2022); count or score.
`SelfEfficacy` LIWC-22 (Boyd et al., 2022); count or score.
`achieve` LIWC-22 (Boyd et al., 2022); count or score.
`acquire` LIWC-22 (Boyd et al., 2022); count or score.
`curiosity` LIWC-22 (Boyd et al., 2022); count or score.
`fulfill` LIWC-22 (Boyd et al., 2022); count or score.
`need` LIWC-22 (Boyd et al., 2022); count or score.
`power` LIWC-22 (Boyd et al., 2022); count or score.
`want` LIWC-22 (Boyd et al., 2022); count or score.

Narrative and reasoning (10)

`Cognition` LIWC-22 (Boyd et al., 2022); count or score.
`Perception` LIWC-22 (Boyd et al., 2022); count or score.
`auditory` LIWC-22 (Boyd et al., 2022); count or score.
`cause` LIWC-22 (Boyd et al., 2022); count or score.
`cogproc` LIWC-22 (Boyd et al., 2022); count or score.
`differ` LIWC-22 (Boyd et al., 2022); count or score.
`memory` LIWC-22 (Boyd et al., 2022); count or score.
`mental` LIWC-22 (Boyd et al., 2022); count or score.
`motion` LIWC-22 (Boyd et al., 2022); count or score.

`visual` LIWC-22 (Boyd et al., 2022); count or score.

Fluency and complexity (17)

`Analytic` LIWC-22 (Boyd et al., 2022); count or score.
`BigWords` LIWC-22 (Boyd et al., 2022); count or score.
`avg_sentence_length` Parser statistic of the text itself; no dictionary; count or score.
`avg_word_length` Parser statistic of the text itself; no dictionary; count or score.
`contractions` Parser statistic of the text itself; no dictionary; count or score.
`exclamations` Parser statistic of the text itself; no dictionary; count or score.
`flesch` Composite of other variables in this set; count or score.
`long_word_count` Parser statistic of the text itself; no dictionary; count or score.
`n_sentences` Parser statistic of the text itself; no dictionary; count or score.
`nonflu` LIWC-22 (Boyd et al., 2022); count or score.
`questions` Parser statistic of the text itself; no dictionary; count or score.
`repeated_bigrams` Parser statistic of the text itself; no dictionary; count or score.
`semantic_diversity` Recovered from the project's earlier feature set; count or score.
`syllables_per_word` Parser statistic of the text itself; no dictionary; count or score.
`top_word_share` Parser statistic of the text itself; no dictionary; share.
`type_token_ratio` Parser statistic of the text itself; no dictionary; ratio.
`word_count` Parser statistic of the text itself; no dictionary; count or score.

Self-presentation and address (10)

`Authentic` LIWC-22 (Boyd et al., 2022); count or score.
`empath_communication_z` Empath (Fast et al., 2016); standardized score.
`empath_family_z` Empath (Fast et al., 2016); standardized score.
`empath_friends_z` Empath (Fast et al., 2016); standardized score.
`female` LIWC-22 (Boyd et al., 2022); count or score.
`i` LIWC-22 (Boyd et al., 2022); count or score.
`male` LIWC-22 (Boyd et al., 2022); count or score.
`they` LIWC-22 (Boyd et al., 2022); count or score.
`we` LIWC-22 (Boyd et al., 2022); count or score.
`you` LIWC-22 (Boyd et al., 2022); count or score.

Substance and time reference (5)

`Stress_avg` LIWC-22 (Boyd et al., 2022); count or score.
`focusfuture` LIWC-22 (Boyd et al., 2022); count or score.
`focuspast` LIWC-22 (Boyd et al., 2022); count or score.
`focuspresent` LIWC-22 (Boyd et al., 2022); count or score.
`work` LIWC-22 (Boyd et al., 2022); count or score.

C.2.3 Structural controls

The 30 predetermined characteristics of the episode and the presenting team, which expand to 67 columns after categorical encoding. They enter as controls and as the baseline prediction, never as a treatment of interest. The regressions carry eleven of these columns, named in Section 4: the season, the position within it, the number of presenters, whether a guest investor was seated, two founder-group indicators, and the five largest product categories. The out-of-sample exercise carries all 67, and Section 6.9 reports the regressions under the full set as well.

`airing_day` *(Numeric)* Day of the month on which the episode aired.

`airing_is_weekend` *(Binary)* One if the episode aired on a Saturday or Sunday.

`airing_month` *(Numeric)* Month in which the episode aired (1–12).

`airing_weekday` *(Categorical)* Weekday on which the episode aired; one indicator per weekday after encoding.

`airing_year` *(Numeric)* Year in which the episode aired.

`episode` *(Numeric)* Episode number within the season.

`founder_group` *(Categorical)* Gender composition of the presenting team: only male, only female, or mixed.

`founder_state` *(Categorical)* Two-letter US state (or country) of the company or founders; a missingness indicator marks pitches where it is not observed.

`guest_gender` *(Categorical)* Gender of the guest investor(s): male, female, mixed, or none.

`has_guest` *(Binary)* One if any guest investor is seated, including regular investors outside their permanent tenure (Mark Cuban in Seasons 1–2, Daniel Lubetzky before Season 16).

`industry` *(Categorical)* NAICS sector of the venture; one indicator per sector after encoding.

`is_barbara_corcoran` *(Binary)* One if Barbara Corcoran is seated on the investor panel.

`is_daniel_lubetzky` *(Binary)* One if Daniel Lubetzky is seated on the investor panel.

`is_daymond_john` *(Binary)* One if Daymond John is seated on the investor panel.

`is_kevin_harrington` *(Binary)* One if Kevin Harrington is seated on the investor panel.

`is_kevin_oleary` *(Binary)* One if Kevin O'Leary is seated on the investor panel.

`is_lori_greiner` *(Binary)* One if Lori Greiner is seated on the investor panel.

`is_mark_cuban` *(Binary)* One if Mark Cuban is seated on the investor panel.

`is_robert_herjavec` *(Binary)* One if Robert Herjavec is seated on the investor panel.

`num_female_founders` *(Numeric)* Number of female presenters.

`num_guests` *(Numeric)* Number of guest investors (anyone seated who is not a permanent investor that season).

`num_male_founders` *(Numeric)* Number of male presenters.

`num_presenters` *(Numeric)* Number of presenters (1–5).

`num_sharks` *(Numeric)* Number of investors seated.

`product_type` *(Categorical)* Product category as listed on the Shark Tank products website; one indicator per category after encoding.

`relationship_std` *(Categorical)* Relationship among the presenters: solo, partners, family, friends, colleagues, or other.

`season` *(Numeric)* Season number.

`season_progress` *(Numeric)* Episode number divided by the number of episodes in the season.

`share_female_founders` *(Numeric)* Share of female presenters (0–1).

`solo_founder` *(Binary)* One if a single founder presents.

C.2.4 Judgment (S_{judgment})

Both models read only the cleaned and anonymized transcript: founder and investor names replaced by placeholders, and the negotiation and closing remarks removed. The prompts are reproduced verbatim. The generation prompt also asks each model to judge whether the pitch ended in a deal; because the transcript excludes the negotiation, this is the model's own assessment and not an observed outcome.

Variable Generation Prompt: Claude

Model Name:	`claude-sonnet-4-6`
Temperature:	0
Max Tokens:	8,912
Batch Size:	50 transcripts per call
Input:	training transcripts only, cleaned and anonymized
Output Format:	JSON

Prompt

```
You are a venture capital investor.

Tasks:
1. Determine decision_from_transcript:
- deal
- no_deal
- unknown
2. Create economically meaningful variables explaining VC funding decisions:
- snake_case
- 1--2 words
- universal

STRICT RULES:
- DO NOT duplicate variables. For example, ask_amount and asking_amount are the same,
implied_valuation and valuation are the same. DO NOT REPEAT any variables more than
once

[Transcripts populated dynamically in batches of 50]
```

Codebook Prompt

```
You are an econometrics researcher.

For each variable:
1. measurement_type:
- binary
- ordinal_0_4
- numeric
2. Provide:
- definition
- methodology
- technique
- provide measurement threshold
```

Variable Generation Prompt: GPT

Model Name:	`gpt-5.4`
Seed:	1234
Batch Size:	50 transcripts per call
Input:	training transcripts only, cleaned and anonymized
Output Format:	JSON

Prompt

```
You are a venture capital investor.

Tasks:
1. Determine decision_from_transcript (deal, no_deal, unknown)
2. Create economically meaningful variables in snake_case that determine VC financing
decisions for early-stage fundraising.
3. It has to be in one or two words, universal, conceptually applies to every pitch,
and academically usable.

STRICT RULES:
1. DO NOT duplicate variables. For example, ask_amount and asking_amount are the same,
implied_valuation and valuation are the same. DO NOT REPEAT any variables more than
once.
```

Codebook Prompt

```
You are an econometrics researcher.

For each variable:
1. Assign measurement_type:
- binary
- ordinal_0_4
- numeric
2. Provide:
- definition
- methodology
- technique
- provide measurement threshold
```

Transcript Scoring Prompt: Claude

Model Name:	`claude-sonnet-4-6`
Temperature:	0.7
Max Tokens:	4,096
Input:	every transcript (training and test), cleaned and anonymized; scored once
Output Format:	JSON

Prompt

```
You are a venture capital researcher.
```

```
Score the pitch transcript strictly using these definitions.

RULES:
- Return ONLY valid JSON
- No explanations
- No markdown
- No text outside JSON

SCORING:
- binary → 0 or 1
- ordinal_0_4 → integer 0--4
- numeric → numbers only


   Do NOT return TRUE/FALSE
Do NOT return text explanations

VARIABLE DEFINITIONS:
[variable_instructions]
```

Transcript Scoring Prompt: GPT

Model Name:	gpt-5.4
Seed:	1234
Input:	every transcript (training and test), cleaned and anonymized; scored once
Output Format:	JSON

Prompt

```
You are a venture capital researcher.  Score the pitch transcript strictly using these
definitions.
Return ONLY structured JSON.

For binary variables return 1 or 0.
For ordinal_0_4 variables return integers 0--4.
For numeric variables, return numbers only.
Do not return TRUE/FALSE.
Do not put anything in text.
do NOT return text explanations
do NOT return markdown
do NOT return anything outside JSON

VARIABLE DEFINITIONS:
[variable_instructions]
```

Judgment variables list (S_{judgment})

The 145 variables the judgment index is built on: 45 proposed by Claude and 116 by GPT, of which 16 were proposed by both. Every variable listed here enters the index, and every variable in the index is listed here. Two further variables were proposed and scored by GPT but have no variation among the training pitches on which the preprocessing is fitted, so they are dropped there and are not listed: cash conversion, which is zero for every pitch it

scores, and the stockout measure, which is zero for every training pitch. Definitions are reproduced verbatim from the codebooks the models generated. Several refer to data sources (financial statements, surveys, media databases) that were not available; the models scored every variable from the cleaned transcript alone. Where both models proposed a variable, the Claude definition is shown.

`average_order` *(Numeric)* Average monetary value of a customer order or transaction. Compute total sales revenue divided by total number of orders over a defined period. Use accounting or transaction data to calculate mean order value; if only stated in the pitch, record the reported figure. Threshold: Greater than or equal to 0 monetary units; undefined if order count equals 0 unless set to missing. Proposed by: GPT.

`b2b_focus` *(Binary)* Indicator of whether the startup's primary revenue model targets business clients (B2B) as opposed to individual consumers (B2C) Coded from pitch descriptions, business model disclosures, and customer segment classifications; dual-coder agreement required with Cohen's Kappa > 0.80 Binary dummy variable (1 = B2B, 0 = B2C or mixed); used as regressor in logit/probit models and as interaction term with deal outcome variables Threshold: 1 if $\geq 70\%$ of stated revenue or customer base derives from business clients; 0 otherwise. Proposed by: Claude.

`backlog` *(Numeric)* Total value or count of confirmed customer orders not yet fulfilled at the time of observation. Measure the outstanding unfulfilled order pipeline using either monetary value or number of orders, keeping units consistent across cases. Extract from firm operational records or entrepreneur statements; standardize as either value-based or count-based metric. Threshold: 0 indicates no backlog; positive values indicate unfulfilled confirmed demand. Proposed by: GPT.

`backorders` *(Numeric)* Number or value of orders delayed because inventory or production capacity is insufficient. Count all orders awaiting fulfillment due to stockouts or production delays, or sum their monetary value. Use inventory and order management records; if disclosed qualitatively in the pitch, code the stated amount when available. Threshold: 0 indicates no backorders; positive values indicate pending delayed orders. Proposed by: GPT.

`barriers_entry` *(Ordinal (0–4))* Degree to which structural factors protect the venture from new competitors entering the market. Code the strength of entry barriers on a 0 to 4 scale based on evidence such as patents, regulation, capital intensity, proprietary technology, network effects, or exclusive partnerships. Content analysis of pitch materials, business descriptions, and supporting documents using a predefined coding rubric. Threshold: 0=no meaningful barriers; 1=weak barriers; 2=moderate barriers; 3=strong barriers; 4=very strong/multiple defensible barriers. Proposed by: GPT.

`brand_strength` *(Ordinal (0–4))* Extent to which the venture has established brand recognition, credibility, and customer attachment. Assess brand strength on a 0 to 4 scale using indicators such as awareness, media presence, repeat customers, endorsements, market reputation, and distinct brand identity. Expert coding from pitch content, market signals, customer traction, and third-party recognition using a standardized rubric. Threshold: 0=no observable brand strength; 1=minimal recognition; 2=emerging brand; 3=strong brand presence; 4=very strong/well-established brand. Proposed by: GPT.

`branding` *(Ordinal (0–4))* Quality and development of the venture's branding strategy and brand presentation. Code branding on a 0 to 4 scale based on clarity of brand identity, visual consistency, messaging, differentiation, and professionalism. Content analysis of pitch decks, packaging, website, promotional material, and verbal presentation with a structured coding guide. Threshold: 0=no evident branding; 1=basic branding; 2=moderate branding development; 3=strong and coherent branding; 4=highly polished and strategically differentiated branding. Proposed by: GPT.

`business_model` *(Ordinal (0–4))* Clarity, viability, and scalability of the venture's mechanism for creating, delivering, and capturing value. Rate the business model on a 0 to 4 scale using evidence on revenue logic, customer segment fit, pricing, cost structure, unit economics, and scalability. Expert assessment based on pitch content and supporting business information using a predefined ordinal rubric. Threshold: 0=no clear business model; 1=poorly defined or weak model; 2=moderately defined/viable model; 3=strong and credible model; 4=highly compelling, scalable, and well-substantiated model. Proposed by: GPT.

`cap_table` *(Ordinal (0–4))* Capitalization table complexity or concentration score reflecting the ownership structure of the firm. Assign a score from 0 to 4 based on the number of shareholder classes, number of investors, and presence of complex instruments such as SAFEs, convertible notes, options, or preferred share layers. Constructed scoring index from legal and financing records. Threshold: 0=very simple founder-only or highly concentrated ownership; 1=simple with few holders/instruments; 2=moderate complexity; 3=high complexity with multiple investor classes/instruments; 4=very high complexity with extensive dilution layers and many security types. Proposed by: GPT.

`capacity` *(Numeric)* The maximum output, service volume, or operational throughput a firm can sustain over a specified period under normal conditions. Measure using operational limits such as units producible, service hours deliverable, or transactions processable per day, week, month, or year. Operational capacity estimation from production records, staffing models, system throughput logs, or engineering constraints. Threshold: capacity > 0 in firm-specific units per period; compare actual output to capacity for utilization, where utilization = actual_output / capacity. Proposed by: GPT.

`category` *(Ordinal (0–4))* Ordinal classification of the startup's industry sector ranked by average historical investor deal rate and sector growth potential, from lowest (0) to highest (4) investment attractiveness Industry sectors mapped to ordinal scale using historical deal closure rates, VC sector preference surveys, and NAICS/SIC classifications; validated against Crunchbase sector benchmarks Ordinal variable treated as ordered categorical; ordered logit model applied; alternatively dummy-coded into sector fixed effects for robustness checks Threshold: 0 = Declining/low-growth sectors (e.g., traditional retail); 1 = Stable low-tech sectors; 2 = Moderate-growth sectors (e.g., consumer goods); 3 = High-growth sectors (e.g., SaaS, fintech); 4 = Emerging/disruptive sectors (e.g., AI, biotech). Proposed by: Claude.

`category_growth` *(Numeric)* The growth rate of the firm's product or market category over a specified time horizon. Compute as the percentage change in category sales, users, volume, or market size from one period to the next. Market-growth estimation using industry reports, syndicated data, or internal market analytics. Threshold: category_growth = ((category_metric_t - category_metric_t-1) / category_metric_t-1) * 100; positive values indicate expansion, negative values indicate contraction. Proposed by: GPT.

`celebrity_endorsement` *(Binary)* Indicator of whether the startup has secured a formal endorsement, partnership, or brand ambassadorship from a recognized public figure or celebrity prior to or during the pitch Coded from pitch content, marketing materials, and public records; celebrity defined as individual with verified social media following > 1,000,000 or mainstream media recognition Binary dummy variable (1 = celebrity endorsement present, 0 = absent); used in logit models to assess signaling effect on deal probability and valuation premium Threshold: 1 if at least one verified celebrity endorsement with documented contractual or public agreement exists; 0 if no endorsement or only micro-influencer partnerships present. Proposed by: Claude.

`channel_fit` *(Ordinal (0–4))* Degree to which a firm's chosen sales and distribution channels align with its product, target customers, and go-to-market model. Construct a 5-point rubric based on alignment between customer preferences, product complexity, service needs, and channel economics. Score from structured managerial assessment or expert coding. Survey/interview-based scoring, expert evaluation, or composite index using indicators such as channel-customer match, cost efficiency, and coverage adequacy. Threshold: 0=no fit; 1=weak fit; 2=moderate fit; 3=good fit; 4=excellent fit. Proposed by: GPT.

`channel_mix` *(Ordinal (0–4))* Extent of diversity and balance in the set of channels used by the firm, such as direct, indirect, online, and partner channels. Assess the breadth and strategic balance of active channels relative to industry norms and firm objectives using a 5-point rubric. Categorical scoring from channel portfolio data or managerial survey, considering number of channels, complementarity, and dependence concentration. Threshold: 0=single/inadequate channel; 1=very limited mix; 2=moderate mix; 3=well-diversified mix; 4=highly optimized multi-channel mix. Proposed by: GPT.

`channel_strategy` *(Ordinal (0–4))* Quality and clarity of the firm's strategic approach to selecting, coordinating, and managing sales/distribution channels. Evaluate formalization, coherence, monitoring, and adaptability of channel planning using a standardized 5-level rubric. Document review, executive survey, or expert rating based on presence of articulated strategy, governance, incentives, and performance tracking. Threshold: 0=no channel strategy; 1=ad hoc strategy; 2=basic strategy; 3=defined and implemented strategy; 4=highly integrated and adaptive strategy. Proposed by: GPT.

`churn` *(Binary)* Indicator of whether a customer, employee, or account discontinued the relationship during the observation period. Code status at the end of a defined period based on contractual termination, inactivity threshold, or account closure. Administrative record linkage, CRM/event-history coding, or panel status tracking. Threshold: 0=retained/active; 1=churned/terminated. Proposed by: GPT.

`claim_risk` *(Ordinal (0–4))* Level of expected probability and severity of future claims associated with a customer, policy, product, or portfolio segment. Classify observed or predicted claim exposure into five ordered categories using historical claims frequency, severity, and underwriting indicators. Actuarial scoring, predictive risk model, or expert underwriting assessment mapped to ordinal bands. Threshold: 0=very low risk; 1=low risk; 2=moderate risk; 3=high risk; 4=very high risk. Proposed by: GPT.

`co_founder` *(Binary)* Indicator of whether the startup was founded by more than one individual, reflecting team-based founding structure versus solo entrepreneurship Coded from pitch disclosures, company registration documents,

and cap table records; co-founder status requires equity ownership and founding-period involvement Binary dummy variable (1 = two or more co-founders, 0 = solo founder); included in probit/logit models as human capital and team diversity signal Threshold: 1 if ≥ 2 individuals hold founding equity stakes established at company inception; 0 if single founder or if additional partners joined post-founding. Proposed by: Claude.

`coachability` *(Ordinal (0–4))* Extent to which an individual is receptive to feedback, willing to learn, and able to improve behavior or performance through guidance. Use behavioral assessment across dimensions such as openness to feedback, responsiveness, and implementation of advice, aggregated into a 5-point score. Supervisor rating, psychometric/behavioral rubric, or structured interview assessment. Threshold: 0=not coachable; 1=minimally coachable; 2=moderately coachable; 3=highly coachable; 4=exceptionally coachable. Proposed by: GPT.

`commitment` *(Ordinal (0–4))* Level of dedication, persistence, and willingness to invest effort toward organizational, contractual, or strategic objectives. Measure through validated attitudinal or behavioral indicators and map average score to a 5-category ordinal scale. Likert-based survey index, manager assessment, or behavioral proxy scoring such as attendance, follow-through, and discretionary effort. Threshold: 0=very low commitment; 1=low commitment; 2=moderate commitment; 3=high commitment; 4=very high commitment. Proposed by: GPT.

`competition` *(Ordinal (0–4))* Perceived or observed degree of rivalry faced by the firm in its product or market space. Rate the overall competitive pressure using factors such as number of rivals, price pressure, differentiation, and entry threat on a 5-point scale. Managerial survey, industry analysis, or expert coding from market reports. Threshold: 0=no meaningful competition; 1=low competition; 2=moderate competition; 3=high competition; 4=very intense competition. Proposed by: GPT.

`competitive_intensity` *(Ordinal (0–4))* Strength of active market rivalry reflected in aggressiveness of competitors, frequency of competitive actions, and pressure on margins or market share. Aggregate indicators of price wars, promotional activity, entry/exit, innovation races, and customer switching pressure into five ordered categories. Survey-based scale, market intelligence coding, or composite index transformed into ordinal bands. Threshold: 0=very low intensity; 1=low intensity; 2=moderate intensity; 3=high intensity; 4=very high intensity. Proposed by: GPT.

`concentration` *(Numeric)* Degree to which market share, sales, customers, or suppliers are concentrated among a small number of entities. Calculate a continuous concentration metric such as the Herfindahl-Hirschman Index, concentration ratio, or share of top entities within the relevant market or portfolio. Quantitative computation from market share or portfolio share data. Threshold: Continuous measure; for HHI example: below 1500=unconcentrated, 1500-2499=moderately concentrated, 2500 and above=highly concentrated. Proposed by: GPT.

`conversion_rate` *(Ordinal (0–4))* The effectiveness with which prospects or leads are converted into paying customers or completed target actions. Assess the observed ratio of conversions to qualified leads, visitors, or trial users over a recent and representative period, using startup reporting, analytics dashboards, or sales funnel data. Score based on empirical funnel performance relative to stage-appropriate expectations and consistency of conversion across periods. Threshold: 0=no measurable conversion data or negligible conversion; 1=low conversion materially below benchmark; 2=moderate conversion near lower acceptable range; 3=strong conversion above average benchmark; 4=excellent conversion consistently high and validated by reliable data. Proposed by: GPT.

`cost_structure` *(Ordinal (0–4))* The degree to which the firm's operating and production costs are efficient, scalable, and aligned with its revenue model. Review fixed versus variable cost mix, burn profile, gross margin structure, operating leverage, and evidence of unit economics from financial statements or management accounts. Score according to efficiency, predictability, and scalability of costs relative to revenue generation and business model maturity. Threshold: 0=cost structure unsustainable or opaque; 1=highly inefficient with poor alignment to revenue; 2=acceptable but with material cost pressure; 3=efficient and reasonably scalable; 4=highly efficient, scalable, and strongly supportive of profitability. Proposed by: GPT.

`credibility` *(Ordinal (0–4))* The extent to which the venture is perceived as trustworthy and legitimate by investors, customers, partners, and the market. Evaluate presence of reputable founders, advisors, investors, partnerships, certifications, media coverage, customer references, and consistency between claims and verifiable evidence. Score based on the strength and external validation of reputational signals and third-party endorsements. Threshold: 0=no credible signals or serious trust concerns; 1=weak credibility with limited validation; 2=some credible indicators but mixed evidence; 3=strong credibility supported by multiple external signals; 4=exceptional credibility with highly reputable endorsements and clear market trust. Proposed by: GPT.

`crowdfunding` *(Binary)* Indicator of whether the startup has previously raised capital through a public crowdfunding campaign via platforms such as Kickstarter, Indiegogo, or equity crowdfunding portals Coded from pitch disclosures, platform records, and SEC Regulation CF filings; distinguishes between reward-based and equity-based crowdfunding Binary dummy variable (1 = prior crowdfunding campaign completed, 0 = no crowdfunding history); used as market

validation proxy in deal outcome regression models Threshold: 1 if at least one crowdfunding campaign reached $\geq$ 50% of stated funding goal or raised $\geq$ $10,000; 0 if no campaign or campaign failed to reach 50% of goal. Proposed by: Claude.

`customer_acquisition` *(Ordinal (0–4))* Ordinal measure of the startup's customer acquisition capability and traction, ranked from no demonstrated acquisition (0) to highly scalable and proven acquisition engine (4) Scored based on combination of customer count, growth rate, CAC metrics, and acquisition channel diversity disclosed in pitch; rubric applied consistently across observations Ordinal variable modeled via ordered logit/probit; Spearman rank correlation used for bivariate analysis; sensitivity analysis with OLS treating variable as continuous Threshold: 0 = No customers or pre-revenue; 1 = Early adopters only (1–50 customers, no repeatable process); 2 = Defined acquisition channel with moderate traction (51–500 customers); 3 = Scalable acquisition with strong growth metrics (501–5,000 customers or MoM growth > 15%); 4 = Proven scalable engine with demonstrated unit economics and > 5,000 customers or viral/network-driven growth. Proposed by: Claude and GPT.

`customer_concentration` *(Ordinal (0–4))* The degree to which revenue depends on a small number of customers. Calculate the share of total revenue attributable to the top 1, top 3, or top 5 customers using recent revenue data. Score inversely with concentration risk, where more diversified customer revenue receives a higher score. Threshold: 0=extreme concentration, top customer >70% of revenue; 1=very high concentration, top customer 40-70% or top 3 >80%; 2=moderate concentration, top customer 20-40% or top 3 50-80%; 3=low concentration, top customer 10-20% and top 3 <50%; 4=high diversification, top customer <10% and no material customer dependency. Proposed by: GPT.

`customer_focus` *(Ordinal (0–4))* The extent to which the venture is oriented around understanding and serving target customer needs. Assess customer research practices, feedback loops, product iteration based on user input, support quality, retention efforts, and clarity of customer segmentation. Score based on evidence of systematic customer-centric decision-making and integration of customer insights into operations and product development. Threshold: 0=no clear customer focus; 1=minimal attention to customer needs; 2=basic customer understanding with irregular feedback use; 3=strong customer focus with regular feedback integration; 4=deeply customer-centric organization with continuous learning and strong customer alignment. Proposed by: GPT.

`customer_traction` *(Ordinal (0–4))* A composite measure of a startup's demonstrated ability to attract, retain, and grow its customer base, reflecting market validation and revenue momentum over a defined observation period Composite index constructed from four sub-indicators: (1) Month-over-Month (MoM) customer growth rate, (2) Monthly Recurring Revenue (MRR) growth, (3) Net Promoter Score (NPS), and (4) Customer Retention Rate. Each sub-indicator is scored 0–4 and aggregated using equal-weighted averaging, then rounded to nearest integer Weighted composite scoring using Likert-type aggregation. Sub-indicators normalized to 0–4 scale using min-max normalization. Cronbach's alpha used to validate internal consistency of composite index (target $\alpha > 0.70$). Final score derived via arithmetic mean of sub-scores Threshold: 0 = No measurable traction (0% MoM growth, no paying customers); 1 = Early traction (1–5% MoM growth, < 50 customers); 2 = Moderate traction (6–15% MoM growth, 50–500 customers); 3 = Strong traction (16–30% MoM growth, 500–5,000 customers); 4 = Exceptional traction (> 30% MoM growth, > 5,000 customers or significant MRR). Proposed by: Claude and GPT.

`customer_validation` *(Ordinal (0–4))* The extent to which real customers have confirmed demand for the product or service through purchasing, usage, pilots, testimonials, or renewals. Collect evidence from paid contracts, pilot outcomes, letters of intent, user engagement, testimonials, churn behavior, and independent customer references. Score based on the strength and quality of external proof that customers value the offering. Threshold: 0=no customer validation; 1=anecdotal or weak validation only; 2=some validation through pilots or limited paid usage; 3=strong validation through paying customers and positive evidence; 4=robust validation through repeat purchasing, renewals, references, and broad customer proof. Proposed by: GPT.

`differentiation` *(Ordinal (0–4))* The extent to which the venture's product, service, technology, or business model is distinct from competitors in ways valued by customers. Compare offering against competitors on features, pricing, technology, intellectual property, user experience, switching costs, and customer-perceived uniqueness. Score based on clarity, defensibility, and market relevance of competitive distinctiveness. Threshold: 0=no meaningful differentiation; 1=weak differentiation easily replicated; 2=moderate differentiation with some customer relevance; 3=strong differentiation that is noticeable and valuable; 4=highly distinctive and defensible differentiation with clear competitive advantage. Proposed by: GPT.

`distribution` *(Ordinal (0–4))* Extent and quality of the firm's ability to deliver, sell, and scale its product through channels, partnerships, or direct market access. Score based on evidence of distribution infrastructure, channel partnerships, customer reach, repeatable sales processes, and scalability of go-to-market operations. Assign ordered score from 0 to 4 using qualitative assessment of market access and channel maturity. Threshold: 0=no identifiable distri-

bution capability; 1=limited or ad hoc channels; 2=some established channels but weak scale; 3=strong distribution with repeatable access; 4=highly developed, scalable, and defensible distribution network. Proposed by: GPT.

`distribution_channel` *(Ordinal (0–4))* An ordinal measure of the effectiveness, scalability, and diversification of a startup's channels through which it delivers its product or service to end customers, including direct, indirect, digital, and partnership-based channels Evaluated across four dimensions: (1) channel diversity (number of active channels), (2) channel scalability (ability to grow without proportional cost increase), (3) channel efficiency (conversion rates and CAC per channel), and (4) channel exclusivity or competitive advantage. Each dimension scored 0–4 and aggregated into a composite ordinal score Structured scoring rubric applied by trained evaluators using standardized questionnaire. Channel mapping conducted via business model canvas analysis. Scalability assessed using channel-specific CAC trends over 6–12 months. Inter-rater reliability validated using Cohen's Kappa (target $\kappa > 0.70$) Threshold: 0 = No defined distribution channel or channel is non-functional; 1 = Single, unproven channel with limited reach and high CAC; 2 = Single established channel with moderate scalability and average efficiency; 3 = Multiple channels with demonstrated scalability and below-average CAC; 4 = Highly diversified, scalable, and proprietary distribution network with strong competitive moat. Proposed by: Claude.

`domain_expertise` *(Ordinal (0–4))* Level of specialized industry, technical, or market knowledge held by the founding or management team relevant to the firm's core business. Score using founders' prior work experience, industry tenure, technical credentials, previous startup exposure, and demonstrated knowledge of the target market. Assign ordered score from 0 to 4 based on documented relevance and depth of experience. Threshold: 0=no relevant expertise; 1=limited familiarity; 2=moderate relevant experience; 3=strong domain knowledge with substantial experience; 4=exceptional expertise with deep track record and recognized credibility. Proposed by: GPT.

`eco_friendly` *(Binary)* A binary indicator denoting whether a startup's core business model, operations, or primary product/service demonstrably reduces environmental impact, promotes sustainability, or aligns with recognized environmental standards and frameworks Assessment based on review of: (1) alignment with UN Sustainable Development Goals (SDGs), particularly SDGs 7, 12, 13, 14, 15; (2) possession of recognized environmental certifications (e.g., B-Corp, ISO 14001, Carbon Neutral certification); (3) documented environmental impact metrics (carbon footprint reduction, waste reduction, renewable energy usage); and (4) ESG disclosure reports Binary classification using a structured checklist of minimum qualifying criteria. At least two of four assessment dimensions must be satisfied for positive classification. Document verification and third-party certification validation applied. Greenwashing risk mitigated through cross-referencing claims with measurable KPIs Threshold: 1 = Eco-friendly (meets minimum 2 of 4 criteria: SDG alignment, environmental certification, measurable impact metrics, ESG reporting); 0 = Not eco-friendly (meets fewer than 2 criteria or environmental claims are unsubstantiated). Proposed by: Claude.

`equity_offer` *(Binary)* Indicator of whether the firm is offering equity securities in the financing round. Code from offering documents, fundraising materials, or transaction terms identifying whether investors receive ownership shares or stock-linked instruments. Assign 1 if the financing includes an equity offer; assign 0 otherwise. Threshold: 0=no equity offer; 1=equity offer present. Proposed by: GPT.

`execution` *(Ordinal (0–4))* Observed capability of the team or firm to deliver milestones, operate effectively, and translate plans into measurable outcomes. Score based on product delivery, milestone attainment, operational discipline, customer acquisition progress, and consistency of implementation relative to stated plans. Assign ordered score from 0 to 4 using qualitative and performance-based evidence. Threshold: 0=no evidence of execution; 1=weak or inconsistent execution; 2=adequate execution with mixed results; 3=strong execution with clear milestone delivery; 4=exceptional execution with repeated, high-quality performance. Proposed by: GPT.

`execution_risk` *(Ordinal (0–4))* Degree of risk that the firm will fail to implement its business plan, launch product, scale operations, or meet key milestones. Score based on complexity of product development, operational dependence, team capability gaps, resource constraints, and historical delivery uncertainty. Assign ordered score from 0 to 4 reflecting increasing risk exposure. Threshold: 0=very low execution risk; 1=low risk; 2=moderate risk; 3=high risk; 4=very high execution risk. Proposed by: GPT.

`exit_potential` *(Ordinal (0–4))* Likelihood and attractiveness of a future liquidity event such as acquisition, merger, or public offering. Score using market size, growth prospects, comparables, strategic acquirer interest, scalability, defensibility, and valuation upside. Assign ordered score from 0 to 4 based on expected probability and magnitude of exit outcomes. Threshold: 0=no credible exit path; 1=limited exit potential; 2=moderate exit potential; 3=strong exit potential; 4=very strong exit potential with multiple plausible high-value exit routes. Proposed by: GPT.

`focus` *(Ordinal (0–4))* Degree to which the firm maintains a clear strategic, product, and market concentration without excessive dispersion of effort. Score based on clarity of value proposition, concentration on core product/customer segment, prioritization discipline, and absence of distracting side initiatives. Assign ordered score from 0 to 4 using qualitative assessment of strategic coherence and operational concentration. Threshold: 0=no clear focus; 1=dif-

fuse strategy with major distractions; 2=some focus but notable fragmentation; 3=strong focus on core priorities; 4=exceptionally clear and disciplined strategic focus. Proposed by: GPT.

`founder_commitment` *(Ordinal (0–4))* Degree to which the founder demonstrates sustained dedication of time, effort, and personal investment to the venture. Code from founder interviews, investor memos, application materials, and behavioral evidence such as prior persistence, personal financial commitment, and duration of involvement. Construct a 5-point rubric based on observable commitment indicators and assign the closest category. Threshold: 0=no observable commitment beyond idea stage; 1=limited/part-time commitment; 2=moderate commitment with clear ongoing involvement; 3=high commitment with substantial time and resource dedication; 4=exceptional commitment with full dedication, strong persistence, and significant personal stake. Proposed by: GPT.

`founder_count` *(Numeric)* The total number of active co-founders involved in the founding and current operation of the startup, defined as individuals who hold a founding title, possess equity stake from inception, and are actively engaged in the business at the time of evaluation Determined through review of incorporation documents, cap table, and founder agreements. Active engagement verified through role confirmation in organizational structure. Advisors, early employees with equity, and inactive founders are excluded from the count. Verified against LinkedIn profiles and company registry filings Integer count derived from legal documentation review and cross-referenced with cap table analysis. Descriptive statistics applied across portfolio for benchmarking. Optimal founder count assessed against empirical research benchmarks (Wasserman, 2012; Kauffman Foundation studies). Count validated through structured founder interviews Threshold: 1 founder: solo founder risk flag; 2–3 founders: optimal range per empirical research (highest success correlation); 4 founders: acceptable with clear role delineation; 5+ founders: elevated governance risk, equity fragmentation concern. Continuous integer variable used in regression models; categorical benchmarks applied for qualitative assessment. Proposed by: Claude.

`founder_experience` *(Ordinal (0–4))* An ordinal composite measure of the founding team's relevant professional experience, domain expertise, prior entrepreneurial history, and demonstrated leadership capability directly applicable to the startup's industry and business model Assessed across five dimensions: (1) years of relevant industry experience, (2) prior startup founding experience (serial entrepreneurship), (3) prior successful exit history (IPO, acquisition), (4) domain-specific technical or functional expertise, and (5) relevant academic credentials or professional certifications. Each dimension scored and aggregated into a composite ordinal score Structured scoring rubric applied through standardized founder assessment questionnaire and CV/LinkedIn verification. Composite score derived via weighted average: prior exits (30%), industry experience (25%), domain expertise (25%), serial entrepreneurship (10%), academic credentials (10%). Weights derived from meta-analysis of venture capital due diligence literature. Inter-rater reliability assessed via Cohen's Kappa (target $\kappa > 0.75$) Threshold: 0 = No relevant experience (first-time founders, no industry background, no relevant credentials); 1 = Limited experience ($<$ 3 years industry experience, no prior startup experience); 2 = Moderate experience (3–7 years industry experience, 1 prior startup with no exit); 3 = Strong experience (7–15 years industry experience, multiple startups, or 1 successful exit); 4 = Exceptional experience (15+ years industry experience, serial entrepreneur with multiple successful exits or IPO history). Proposed by: Claude and GPT.

`founder_fit` *(Ordinal (0–4))* Alignment between the founder's background, skills, motivation, and the specific problem, market, and business model of the venture. Evaluate founder-market fit using educational background, career history, domain expertise, lived experience, and stated motivation relative to the venture opportunity. Apply a 5-point expert coding rubric based on degree of strategic and experiential alignment. Threshold: 0=no discernible fit; 1=weak fit with limited relevance; 2=moderate fit with some matching skills or experience; 3=strong fit with clear domain and execution alignment; 4=exceptional founder-market fit with unique credibility, insight, and capability. Proposed by: GPT.

`founder_invested` *(Binary)* Indicates whether the founder(s) have personally invested their own capital into the venture, signaling financial commitment and skin-in-the-game alignment Survey or financial disclosure data collection confirming personal capital contribution by at least one founding member; verified through cap table or financial statements Dummy variable coding: 1 if founder has made a personal financial investment into the company, 0 otherwise; logistic regression or probit model applicable Threshold: 1 = Founder has personally invested any amount of personal capital; 0 = No personal financial investment made by founder. Proposed by: Claude.

`franchise_model` *(Binary)* Indicates whether the business operates under a franchise business model, where the company licenses its brand, systems, and processes to independent operators Business model classification derived from legal agreements, business registration documents, or structured survey responses identifying franchise licensing arrangements Dummy variable coding: 1 if the business operates as a franchisor or franchisee, 0 if operating under an independent or alternative business model Threshold: 1 = Business operates as a franchise (franchisor or franchisee); 0 = Business does not operate under a franchise model. Proposed by: Claude.

`franchising` *(Binary)* Indicator for whether the venture's business model includes franchising as a primary expansion or operating mechanism. Identify from company descriptions, legal filings, websites, pitch decks, or operational model documentation whether franchising is used. Dummy-variable coding from documented business model classification. Threshold: 0=venture does not franchise; 1=venture uses or plans franchising as a core business model. Proposed by: GPT.

`full_time` *(Binary)* Indicator for whether the founder is working on the venture on a full-time basis. Verify from founder self-report, employment records, application forms, or investor due diligence notes whether the founder's primary occupation is the venture. Dummy-variable coding based on employment status at time of observation. Threshold: 0=founder is part-time or simultaneously employed elsewhere as primary occupation; 1=founder works full-time on the venture. Proposed by: GPT.

`full_time_founders` *(Ordinal (0–4))* The number of founders who are working full-time (primary occupation) on the venture, reflecting founding team commitment and operational capacity Survey-based self-reporting or HR records confirming the count of founders dedicating 100% of professional time to the startup; cross-validated with payroll or equity records Ordinal encoding from 0 to 4; ordered logistic regression or Spearman rank correlation; treated as ordinal due to diminishing marginal returns beyond 4 full-time founders Threshold: 0 = No full-time founders; 1 = One full-time founder; 2 = Two full-time founders; 3 = Three full-time founders; 4 = Four or more full-time founders. Proposed by: Claude.

`go_to_market` *(Ordinal (0–4))* Quality and readiness of the venture's go-to-market strategy, including target customer definition, channel strategy, pricing, and sales execution plan. Review business plans, pitch decks, strategy documents, and commercial traction evidence for specificity and feasibility of market entry approach. Expert rubric scoring on a 5-point scale based on clarity, coherence, channel suitability, and execution readiness. Threshold: 0=no articulated go-to-market strategy; 1=very weak or vague strategy; 2=basic strategy with partial customer/channel definition; 3=well-developed strategy with credible channels and pricing; 4=highly robust, validated, and execution-ready go-to-market plan. Proposed by: GPT.

`governance` *(Ordinal (0–4))* Strength of the venture's governance structure, including decision rights, oversight, board/advisor quality, controls, and accountability mechanisms. Assess governance documents, board composition, advisory structure, reporting routines, and formal control systems. Score governance maturity using a 5-point rubric reflecting formality and effectiveness of oversight arrangements. Threshold: 0=no formal governance; 1=very limited informal governance; 2=basic governance with some defined roles or advisors; 3=strong governance with active board/advisors and reporting; 4=highly developed governance with formal oversight, controls, and strategic accountability. Proposed by: GPT.

`growth` *(Ordinal (0–4))* Categorical assessment of the venture's recent business expansion in sales, users, employees, locations, or other key scale indicators. Use historical operating data over a defined recent period and classify observed expansion into ordered growth categories. Convert continuous growth evidence into a 5-point ordered scale using predefined performance bands. Threshold: 0=no growth or decline; 1=low growth; 2=moderate growth; 3=high growth; 4=exceptional growth relative to prior period and peer context. Proposed by: GPT.

`growth_rate` *(Numeric)* Percentage change in a selected performance metric, typically revenue, users, or sales, over a specified period. Compute from two time-period observations as (Current Period - Prior Period) / Prior Period multiplied by 100, using consistent interval definitions. Continuous percentage-change calculation from panel or time-series business data. Threshold: growth_rate = ((current_value - prior_value) / prior_value) * 100; values below 0 indicate contraction, 0 indicates no growth, and positive values indicate expansion. Proposed by: GPT.

`has_cofounder` *(Binary)* Indicates whether the founding team consists of more than one founder, capturing the presence of a co-founding partnership versus a solo founder structure Derived from cap table records, company registration documents, or founder survey responses identifying multiple individuals holding founding equity stakes Dummy variable coding: 1 if two or more co-founders exist, 0 if solo founder; used in logistic regression, survival analysis, or as a moderating variable in startup performance models Threshold: 1 = Two or more co-founders present; 0 = Solo founder with no co-founding partner. Proposed by: Claude.

`inventory` *(Numeric)* Total monetary value or physical quantity of goods, raw materials, work-in-progress, and finished products held by the firm at a given time. Measure using accounting records or inventory management systems at book value or unit count for the reporting period. Use ERP/accounting data to compute ending inventory, average inventory, or inventory units depending on analytical purpose. Threshold: Continuous numeric measure; threshold not fixed. Common practical cutoffs may be defined contextually, e.g., 0=no inventory, >0=inventory held. Proposed by: GPT.

`inventory_level` *(Numeric)* The total monetary value of goods held in stock by the company at a given point in time, including raw materials, work-in-progress, and finished goods, measured in USD Sourced from balance sheet data

under current assets; calculated using FIFO, LIFO, or weighted average cost methods; normalized by revenue or total assets for cross-company comparability Continuous numeric variable; ratio analysis (Inventory/Revenue or Inventory Turnover = COGS/Inventory); OLS regression with industry controls; log-transformation for skewed distributions Threshold: Continuous scale in USD; Inventory Turnover benchmarks: $<$ 2x = overstocked/slow-moving; 2–6x = moderate; $>$ 6x = lean/efficient; context-dependent by industry (retail vs. SaaS). Proposed by: Claude.

`inventory_risk` *(Ordinal (0–4))* Exposure of the firm to losses arising from excess, obsolete, perishable, volatile, or stockout-prone inventory. Score using indicators such as holding period, obsolescence rate, demand volatility, spoilage risk, and stockout frequency. Apply a 5-point risk rubric combining inventory turnover, write-downs, perishability, and forecast error metrics. Threshold: 0=negligible risk; 1=low risk; 2=moderate risk; 3=high risk; 4=very high/severe inventory risk. Proposed by: GPT.

`inventory_status` *(Ordinal (0–4))* A categorical assessment of the company's inventory health and management efficiency, ranging from critically understocked to severely overstocked, reflecting supply chain and demand alignment Derived from inventory turnover ratios, days inventory outstanding (DIO), and qualitative management assessments; scored using a structured rubric comparing actual vs. optimal inventory levels Ordinal encoding 0–4; ordered logistic regression or proportional odds model; Spearman correlation with financial performance metrics; thresholds calibrated by industry benchmarks Threshold: 0 = Critically understocked (stockouts, lost sales); 1 = Below optimal inventory levels; 2 = Optimal/balanced inventory; 3 = Slightly overstocked (excess carrying costs); 4 = Severely overstocked (obsolescence risk, cash flow strain). Proposed by: Claude.

`ip_risk` *(Ordinal (0–4))* Risk that the firm's intellectual property position is weak, infringed, unenforceable, exposed to litigation, or insufficiently protected. Assess based on patent coverage, freedom-to-operate analysis, litigation exposure, geographic protection, and dependency on unprotected know-how. Use legal/IP due diligence and expert scoring with a 5-point rubric incorporating registration status, infringement exposure, and enforceability. Threshold: 0=no meaningful IP risk; 1=low risk; 2=moderate risk; 3=high risk; 4=very high/critical IP risk. Proposed by: GPT.

`ip_strength` *(Ordinal (0–4))* Strength and defensibility of the firm's intellectual property portfolio, including patents, trademarks, copyrights, trade secrets, and contractual protections. Score using breadth of protection, enforceability, remaining life, geographic coverage, and strategic relevance to revenue generation. Apply a 5-point rubric based on IP portfolio quality, number and scope of filings, legal defensibility, and commercialization relevance. Threshold: 0=no meaningful IP protection; 1=weak IP position; 2=moderate IP strength; 3=strong IP portfolio; 4=very strong/core defensible IP. Proposed by: GPT.

`lead_time` *(Numeric)* Elapsed time between initiation of an order, production step, or customer request and final delivery or completion. Measure as the average, median, or total duration using operational timestamps from procurement, production, or fulfillment systems. Calculate time differences from process logs or ERP data, typically in days or hours, over a defined observation window. Threshold: Continuous numeric measure; threshold not fixed. Common contextual cutoff: 0 indicates immediate availability, higher values indicate longer delays. Proposed by: GPT.

`liability_risk` *(Ordinal (0–4))* Exposure to legal, financial, or regulatory obligations arising from product failure, service harm, contractual breach, or non-compliance. Assess using legal claims history, regulatory exposure, product safety profile, indemnification obligations, and insurance coverage gaps. Use a 5-point expert risk scoring framework informed by legal due diligence, compliance reviews, and historical incident data. Threshold: 0=negligible liability risk; 1=low risk; 2=moderate risk; 3=high risk; 4=very high/severe liability risk. Proposed by: GPT.

`licensing_deal` *(Binary)* Indicates whether the company has entered into a formal licensing agreement to monetize its intellectual property, technology, brand, or proprietary processes by granting usage rights to third parties Identified through legal contract disclosures, SEC filings, press releases, or structured survey responses confirming the existence of at least one active licensing agreement generating royalty or fee income Dummy variable coding: 1 if at least one active licensing deal exists, 0 otherwise; logistic regression, event study analysis, or used as moderating variable in revenue diversification models Threshold: 1 = At least one active licensing agreement in place generating contractual rights/royalties; 0 = No licensing agreements exist. Proposed by: Claude.

`licensing_model` *(Binary)* Indicates whether a business operates under a licensing or franchising model, where intellectual property, brand, or operational systems are licensed to third parties for revenue generation Determined through business registration documents, contractual agreements, and revenue stream analysis to confirm presence or absence of formal licensing arrangements Binary classification coded as 1 if the business derives any revenue from licensing agreements or operates as a licensor/licensee, and 0 if no licensing model exists. Verified through legal documentation and financial statements Threshold: 1 = Licensing model present (any active licensing agreement generating revenue); 0 = No licensing model present. Proposed by: Claude and GPT.

`lifetime_value` *(Numeric)* The projected net present value of the total revenue or profit a business expects to generate from a customer or the entire customer base over the duration of the business relationship, representing long-term economic worth Calculated using discounted cash flow analysis applied to historical customer transaction data, average purchase frequency, retention rates, and gross margin. Customer cohort analysis used to estimate churn rates and average customer lifespan LTV = (Average Purchase Value × Purchase Frequency × Gross Margin × Customer Lifespan) discounted at WACC. Alternatively computed as LTV = ARPU × Gross Margin / Churn Rate for subscription models. Regression-based predictive modeling using historical cohort data Threshold: Continuous variable measured in USD per customer or aggregate USD. Minimum: $0; Segmentation benchmarks: Low LTV < $100; Medium LTV $100–$1,000; High LTV $1,000–$10,000; Very High LTV > $10,000. Proposed by: Claude and GPT.

`ltv` *(Numeric)* Abbreviation for lifetime value; estimated total expected value contributed by a customer across the relationship horizon. Measure identically to lifetime_value using expected revenues, margins, retention patterns, and relevant costs. Estimate with CLV/LTV modeling using cohort methods, discounted expected cash flows, or probabilistic customer models. Threshold: Continuous numeric measure; threshold not fixed. Common practical cutoff: <=0 indicates non-profitable customer value, >0 indicates positive value. Proposed by: GPT.

`manufacturing` *(Binary)* Indicates whether the firm materially engages in in-house or owned manufacturing of its core products. Review company filings, annual reports, investor presentations, and official website disclosures to determine if production is performed internally rather than fully outsourced. Code 1 if the company owns, operates, or directly controls manufacturing facilities for core products; code 0 otherwise. Threshold: 1 = in-house/owned manufacturing present; 0 = no in-house manufacturing or fully outsourced production. Proposed by: GPT.

`manufacturing_location` *(Binary)* Indicates whether a business conducts its primary manufacturing or production operations domestically (within the home country) versus internationally (offshore or nearshore), capturing the geographic origin of production Determined through supply chain documentation, import/export records, facility registration, and cost-of-goods-sold analysis to identify where the majority of physical production or assembly occurs Binary classification based on the location where greater than 50% of production value is added. Verified through customs documentation, supplier contracts, and facility audits. For mixed models, classification based on primary production site by volume Threshold: 1 = Domestic manufacturing (>50% of production value added within home country); 0 = International/offshore manufacturing (>50% of production value added outside home country). Proposed by: Claude.

`market_growth` *(Ordinal (0–4))* Rate at which the firm's primary addressable market is expanding over time. Estimate compound annual growth rate (CAGR) of the relevant market using industry reports, market research databases, or regulatory/industry publications over a 3- to 5-year horizon. Map market CAGR into a 0-4 ordered scale based on predefined growth bands. Threshold: 0 = negative growth; 1 = 0% to <5% CAGR; 2 = 5% to <10% CAGR; 3 = 10% to <20% CAGR; 4 = >=20% CAGR. Proposed by: GPT.

`market_size` *(Ordinal (0–4))* A categorical assessment of the total addressable market (TAM) available to the business, representing the overall revenue opportunity if the business were to achieve 100% market share within its defined industry and geographic scope Estimated using top-down industry reports (IBISWorld, Statista, Grand View Research), bottom-up customer segmentation analysis, and serviceable addressable market calculations. Market size validated against competitor revenue benchmarks and industry association data Ordinal scoring based on TAM estimation using combination of top-down and bottom-up approaches. TAM calculated as: Total potential customers × Average annual spend per customer. Scored on 0–4 Likert-type scale based on logarithmic revenue brackets Threshold: 0 = Micro market (TAM < $1M); 1 = Small market (TAM $1M–$10M); 2 = Medium market (TAM $10M–$100M); 3 = Large market (TAM $100M–$1B); 4 = Massive market (TAM > $1B). Proposed by: Claude and GPT.

`media_coverage` *(Ordinal (0–4))* A measure of the extent and quality of earned media attention a business has received across print, digital, broadcast, and social media channels, reflecting public visibility, brand awareness, and reputational reach Assessed through media monitoring tools (Meltwater, Cision, Google News), social listening platforms, and manual audit of press mentions. Coverage scored based on outlet reach, sentiment, frequency of mentions, and prominence of placement Composite ordinal scoring using media audit methodology: (1) Count of unique media mentions, (2) Weighted by outlet audience size/domain authority, (3) Adjusted for sentiment (positive/neutral/negative), (4) Normalized to 0–4 scale using percentile ranking within industry cohort Threshold: 0 = No media coverage (0 mentions); 1 = Minimal coverage (local/niche outlets only, 1–5 mentions); 2 = Moderate coverage (regional outlets or industry press, 6–20 mentions); 3 = Significant coverage (national outlets or major industry publications, 21–100 mentions); 4 = Extensive coverage (major national/international outlets, viral social media, >100 high-reach mentions). Proposed by: Claude.

`mission` *(Binary)* Indicates whether the company has a clearly articulated mission statement or equivalent purpose

declaration. Inspect official company sources such as the corporate website, annual report, ESG report, or founder letters for an explicit mission or purpose statement. Code 1 if a formal mission/purpose statement is publicly stated; code 0 if absent. Threshold: 1 = explicit public mission statement present; 0 = no explicit public mission statement. Proposed by: GPT.

`mission_alignment` *(Ordinal (0–4))* Degree to which the firm's products, operations, and strategic actions are consistent with its stated mission. Compare the stated mission with revenue sources, product portfolio, acquisitions, strategic initiatives, and operating practices using company disclosures and credible third-party sources. Assign an ordered score from 0 to 4 based on the observed consistency between mission and actual business conduct. Threshold: 0 = no evident alignment; 1 = weak alignment; 2 = moderate alignment; 3 = strong alignment; 4 = very strong/full alignment. Proposed by: GPT.

`mission_driven` *(Binary)* Indicates whether a business explicitly incorporates a social, environmental, or community-oriented mission as a core component of its business model and strategic objectives, beyond pure profit maximization Assessed through review of founding documents, mission statements, B-Corp certification status, legal structure (benefit corporation, social enterprise), marketing materials, and ESG reporting. Validated against operational evidence of mission integration in business decisions Binary classification based on structured content analysis of organizational documents and third-party certifications. Requires at least two of the following: (1) explicit mission statement referencing social/environmental impact, (2) legal structure as benefit corp or nonprofit, (3) third-party certification (B-Corp, 1% for the Planet), (4) dedicated impact reporting Threshold: 1 = Mission-driven (explicit social/environmental mission embedded in business model with operational evidence); 0 = Not mission-driven (purely profit-oriented without formal social/environmental mission). Proposed by: Claude.

`mission_strength` *(Ordinal (0–4))* Strength, clarity, and centrality of the company's mission in guiding strategy and external communication. Assess prominence of mission language in company materials, leadership messaging, strategic documents, and evidence of mission-linked decision-making. Score on a 0-4 ordinal scale reflecting how clearly and consistently mission is defined and embedded. Threshold: 0 = absent; 1 = vague/minor; 2 = defined but limited influence; 3 = clear and important; 4 = highly explicit and central to strategy. Proposed by: GPT.

`moat` *(Ordinal (0–4))* Strength of the firm's durable competitive advantage relative to rivals. Evaluate evidence on switching costs, brand power, network effects, cost advantages, proprietary technology, patents, data advantages, regulation, and distribution scale from filings and market analysis. Assign a 0-4 ordinal score based on the breadth and durability of competitive advantage. Threshold: 0 = no identifiable moat; 1 = weak moat; 2 = moderate moat; 3 = strong moat; 4 = very strong/durable moat. Proposed by: GPT.

`months_operating` *(Numeric)* The total number of calendar months a business has been in active operation from its official founding or incorporation date to the current measurement date, serving as a proxy for business maturity, experience, and operational tenure Calculated from official business registration date, incorporation documents, or first revenue date (whichever is earliest and verifiable) to the survey/assessment date. Verified through state/federal business registration records and tax filing history Integer count of complete calendar months elapsed since founding date. Calculated as: (Assessment Date - Founding Date) in months, rounded down to nearest whole month. Used directly in regression models or transformed using square root transformation to address non-linearity in maturity effects Threshold: Continuous integer variable. Minimum: 0 months (pre-launch); Stage benchmarks: Pre-revenue: 0–6 months; Early-stage: 7–24 months; Growth-stage: 25–60 months; Established: 61–120 months; Mature: >120 months. Proposed by: Claude.

`order_value` *(Numeric)* Average monetary value per customer order over a specified period. Collect total gross merchandise value or sales attributable to orders and divide by total number of orders, using company operational metrics or internal transaction data. Calculate order_value = total order-related revenue / number of orders. Threshold: Continuous currency measure; >0 with no fixed ordinal cutoff. Proposed by: GPT.

`outside_capital` *(Binary)* Indicator of whether the firm obtained external financing from sources outside the founding team, such as venture capital, angel investment, bank loans, grants, or strategic investors. Code as 1 if the firm reports any receipt of external capital during the reference period; code as 0 if financing is entirely internal or bootstrapped. Survey item, financing history review, or administrative/funding records verification. Threshold: 0 = no outside capital received; 1 = any outside capital received. Proposed by: GPT.

`ownership` *(Numeric)* Percentage equity ownership held by a specified founder, owner, or stakeholder in the venture. Measure as the share of total outstanding equity attributable to the focal party at the time of observation. Cap table extraction, shareholder registry review, or self-reported equity share validated against company records. Threshold: Continuous measure from 0 to 100 percent ownership. Proposed by: GPT.

`ownership_retained` *(Numeric)* Percentage of original or pre-financing equity ownership that remains with the focal founder or initial owners after investment or restructuring events. Measure the remaining equity share held by the

focal owner(s) at the observation date, or compute retained share relative to initial ownership. Cap table analysis, shareholder agreement review, or founder survey cross-checked with records. Threshold: Continuous measure from 0 to 100 percent retained ownership. Proposed by: GPT.

`partnership_pipeline` *(Numeric)* Number of prospective partnerships under active development but not yet finalized. Count partnership opportunities that are in formal discussion, negotiation, due diligence, or proposal stages during the reference period. CRM/pipeline database extraction, managerial survey, or business development records review. Threshold: Non-negative count: 0, 1, 2, ... active prospective partnerships. Proposed by: GPT.

`partnerships` *(Numeric)* Total number of formalized partnerships maintained by the firm within the specified period. Count executed partnership agreements, alliances, collaborations, or joint ventures that meet the study's inclusion criteria. Contract review, alliance database extraction, or self-reported partnership inventory validated with documentation. Threshold: Non-negative count: 0, 1, 2, ... formal partnerships. Proposed by: GPT.

`patent_protection` *(Binary)* Indicator of whether the firm's innovation or product is protected by at least one patent or active patent-related legal protection. Code as 1 if the firm holds at least one granted patent or enforceable patent-based protection relevant to the focal technology; code as 0 otherwise. Patent portfolio review, legal/IP records verification, or survey response validated through patent databases. Threshold: 0 = no patent protection; 1 = at least one patent-protected asset. Proposed by: GPT.

`patent_status` *(Binary)* Indicates whether a company or product holds at least one active, pending, or granted patent protecting its intellectual property Binary classification based on verified patent filings through USPTO, EPO, or equivalent national patent offices; cross-referenced with company disclosures and legal documentation Logistic regression dummy variable coding; 1 = patent exists (granted or pending), 0 = no patent protection; verified via patent database API queries Threshold: 1 = Has at least one active/pending/granted patent; 0 = No patent protection of any kind. Proposed by: Claude and GPT.

`patents` *(Numeric)* Number of patents associated with the firm, product, or focal innovation. Count granted patents, or if specified by the study, count total patent applications and grants separately under a harmonized rule. Patent database search, IP portfolio review, or firm-reported count validated against official registries. Threshold: Non-negative count: 0, 1, 2, ... patents. Proposed by: GPT.

`pilot_results` *(Ordinal (0–4))* Degree of success demonstrated by pilot tests, beta programs, or trial deployments of the product with early users or customers. Code based on observed pilot performance using evidence such as user adoption, retention, conversion, satisfaction, and achievement of predefined pilot objectives. Ordinal scoring from documented pilot outcomes, customer testimonials, and performance metrics. Threshold: 0=no pilot conducted or no evidence; 1=pilot conducted with poor or inconclusive results; 2=mixed or modestly positive pilot results; 3=strong positive pilot results with clear customer validation; 4=exceptional pilot results with high adoption, strong outcomes, and clear scale potential. Proposed by: GPT.

`pre_sales` *(Binary)* Indicates whether a company has generated any revenue or committed purchase orders prior to full product launch or during pre-launch phase Binary classification derived from financial statements, pitch decks, investor disclosures, and CRM records confirming revenue or binding purchase commitments before official market launch Logistic regression dummy variable; coded from revenue records and signed purchase agreements; 1 = pre-sales revenue or commitments exist, 0 = no pre-sales activity recorded Threshold: 1 = At least one confirmed pre-sale transaction or binding purchase order exists; 0 = Zero pre-sale transactions or commitments. Proposed by: Claude.

`innovation` *(Ordinal (0–4))* Degree to which the firm's product, service, process, or business model is novel and differentiated relative to existing market alternatives. Score based on structured expert assessment of novelty, R&D intensity, uniqueness of value proposition, and evidence of new intellectual or commercial application. Construct a 5-point rubric using product novelty, technological originality, and market differentiation indicators from company documents, patents, and expert review. Threshold: 0=none/minimal innovation; 1=incremental improvement; 2=moderate differentiation; 3=high innovation with clear novelty; 4=breakthrough/disruptive innovation. Proposed by: GPT.

`preorders` *(Ordinal (0–4))* Extent of customer demand demonstrated through advance purchases, reservations, or committed orders before full product launch. Code based on the volume and quality of preorder commitments relative to the firm's target market, production expectations, or campaign goals. Ordinal scoring using preorder counts, deposit-backed reservations, and revenue from advance sales. Threshold: 0=no preorders; 1=very limited preorders or weak demand signal; 2=moderate preorders indicating some market interest; 3=strong preorder volume indicating substantial demand; 4=exceptionally high preorder volume substantially exceeding expectations or targets. Proposed by: GPT.

`pricing` *(Ordinal (0–4))* Assessment of the appropriateness and competitiveness of the venture's pricing strategy

relative to customer value and market benchmarks. Code by comparing stated pricing to competitors, customer willingness to pay, margins, and perceived value proposition. Ordinal scoring based on benchmark analysis, customer feedback, and pricing strategy review. Threshold: 0=pricing absent, unclear, or clearly misaligned with market; 1=weak pricing strategy with major competitiveness issues; 2=adequate pricing strategy with some concerns; 3=strong pricing strategy well aligned with value and market; 4=excellent pricing strategy with clear optimization and competitive advantage. Proposed by: GPT.

`prior_exit` *(Binary)* Indicates whether the founding team or key executives have previously achieved a successful exit event such as acquisition, IPO, or merger in a prior venture Binary classification based on founder background verification through Crunchbase, LinkedIn, SEC filings, and M&A databases confirming prior successful exit events Logistic regression dummy variable; coded from verified founder history; 1 = at least one founder has a documented prior exit, 0 = no prior exit among founding team Threshold: 1 = At least one founding team member has a verified prior exit (acquisition, IPO, or merger); 0 = No founding team member has a prior exit. Proposed by: Claude.

`product_differentiation` *(Ordinal (0–4))* Extent to which the product is distinct from alternatives in features, design, technology, branding, or customer value proposition. Code based on comparison with competing products and assessment of uniqueness that is meaningful to customers. Ordinal scoring using competitor benchmarking, expert review, and customer perception evidence. Threshold: 0=no meaningful differentiation; 1=limited differentiation with mostly commoditized offering; 2=moderate differentiation on some attributes; 3=strong differentiation clearly noticeable in the market; 4=exceptional differentiation with highly unique and defensible value proposition. Proposed by: GPT.

`product_market_fit` *(Ordinal (0–4))* Degree to which the product satisfies a strong market need and is validated by customer demand, usage, and retention. Code based on evidence from customer adoption, repeat usage, retention, net promoter indicators, revenue traction, and qualitative feedback. Ordinal scoring from market validation metrics and customer interviews. Threshold: 0=no evidence of product-market fit; 1=weak fit with limited customer resonance; 2=partial fit with some validated demand; 3=strong fit with clear and repeatable customer demand; 4=exceptional fit with very strong traction, retention, and customer advocacy. Proposed by: GPT.

`product_quality` *(Ordinal (0–4))* Overall standard of the product in terms of reliability, functionality, durability, usability, and customer satisfaction. Code based on defect rates, returns, reviews, expert evaluation, and adherence to performance standards. Ordinal scoring using quality assurance records, customer ratings, and expert assessment. Threshold: 0=very poor quality with major defects or unusability; 1=below-average quality with frequent issues; 2=acceptable quality meeting basic expectations; 3=high quality with few issues and strong customer satisfaction; 4=excellent quality with superior performance and minimal defects. Proposed by: GPT.

`product_stage` *(Ordinal (0–4))* Measures the current development and commercialization stage of the primary product or service on a five-point ordinal scale from concept to scaled market deployment Ordinal scoring based on product development milestones assessed through technical documentation, demo availability, customer traction data, and revenue records following a standardized stage-gate framework Ordered logistic regression; scored by trained evaluators using standardized rubric; inter-rater reliability assessed via Cohen's Kappa; scores averaged across minimum two independent raters Threshold: 0 = Concept/Idea only (no prototype); 1 = Prototype/MVP in development; 2 = Beta/MVP tested with limited users; 3 = Product launched with early revenue; 4 = Scaled product with consistent revenue and market traction. Proposed by: Claude and GPT.

`product_uniqueness` *(Ordinal (0–4))* Measures the degree to which a product or service offers differentiated, novel, or proprietary features that distinguish it from existing market alternatives Ordinal scoring based on competitive landscape analysis, patent review, feature differentiation mapping, and expert panel evaluation assessing novelty, defensibility, and market differentiation Ordered logistic regression; scored by domain expert panel using structured rubric; competitive benchmarking against top 3-5 market alternatives; Likert-style aggregation with normalized scoring Threshold: 0 = No differentiation (commodity product); 1 = Minimal differentiation (minor feature variation); 2 = Moderate differentiation (notable but replicable features); 3 = Strong differentiation (significant proprietary features); 4 = Highly unique (patented, defensible, no direct comparable). Proposed by: Claude.

`purchase_orders` *(Numeric)* Count of formal purchase orders received from customers or buyers within a specified period. Sum all valid purchase orders recorded in procurement or sales systems during the observation window. Administrative count from transaction records. Threshold: 0 = no purchase orders; >=1 indicates at least one purchase order. Proposed by: GPT.

`regulatory_risk` *(Ordinal (0–4))* Level of exposure to adverse effects from laws, regulations, compliance burdens, or policy changes. Assign a score based on assessed severity and likelihood of regulatory constraints affecting operations. Expert risk scoring on a 5-point ordinal scale. Threshold: 0=none, 1=low, 2=moderate, 3=high, 4=very high regulatory risk. Proposed by: GPT.

`repeat_customers` *(Numeric)* The absolute count or percentage of customers who have made more than one purchase or engagement transaction within a defined observation period Derived from CRM systems, transaction databases, or customer account records; customers identified by unique ID; repeat status confirmed by minimum two distinct purchase events within the measurement window OLS regression as continuous variable; expressed as either raw count or percentage of total customer base; cohort analysis used to control for customer tenure; measurement window standardized to 12-month period Threshold: Continuous numeric variable; expressed as count (0 to N) or percentage (0% to 100%); minimum threshold for repeat classification = 2 or more purchases within 12-month window; industry benchmark typically 20-40% repeat rate. Proposed by: Claude.

`repeat_purchase` *(Binary)* Indicates whether a customer has made at least one additional purchase following their initial transaction within a defined observation period Binary classification derived from transaction-level CRM or POS data; each customer assigned binary status based on presence of two or more distinct purchase events within the standardized observation window Logistic regression dummy variable; 1 = customer has made at least one repeat purchase within observation period, 0 = customer has made only one purchase (no repeat); survival analysis used for time-to-repeat modeling Threshold: 1 = Customer has completed 2 or more distinct purchase transactions within the 12-month observation window; 0 = Customer has completed only 1 purchase transaction. Proposed by: Claude and GPT.

`repeat_rate` *(Numeric)* The proportion of total customers who have made more than one purchase within a defined period, expressed as a percentage of the total active customer base Calculated as (number of customers with 2+ purchases / total number of customers) x 100 within a standardized 12-month rolling window; derived from transaction-level CRM or e-commerce platform data OLS regression as continuous percentage variable; beta regression used when variable is bounded between 0 and 1; cohort-controlled analysis to account for customer acquisition timing; winsorized at 5th and 95th percentiles Threshold: Continuous numeric variable expressed as percentage (0% to 100%); 0% = no customers have repeated; 100% = all customers have repeated; industry benchmarks: <20% low retention, 20-40% moderate, 40-60% strong, >60% excellent retention. Proposed by: Claude and GPT.

`resilience` *(Ordinal (0–4))* Capacity of a firm or system to withstand, adapt to, and recover from disruptions. Score based on observed preparedness, continuity capability, recovery speed, and adaptive capacity. Composite expert assessment using a 5-point ordinal rubric. Threshold: 0=none, 1=low, 2=moderate, 3=high, 4=very high resilience. Proposed by: GPT.

`retail_distribution` *(Numeric)* The breadth and depth of a product's availability across retail outlets, measured as the percentage of total possible retail locations or weighted distribution coverage within a defined market Calculated as the ratio of retail outlets carrying the product to the total universe of relevant retail outlets, often weighted by store volume or sales contribution. Data sourced from retail audits, syndicated data providers (e.g., Nielsen, IRI), or internal sales records Weighted Distribution (%) = (Sum of sales volume of stores carrying product / Total sales volume of all stores in market) × 100; alternatively, Numeric Distribution (%) = (Number of stores carrying product / Total number of stores) × 100 Threshold: 0–100%; Low: <25%, Moderate: 25–60%, High: >60%; Full distribution benchmark typically set at ≥85%. Proposed by: Claude.

`retail_doors` *(Numeric)* The total count of individual retail store locations (physical points of sale) where a product or brand is actively sold and stocked, representing the raw unit count of distribution points Aggregated from internal sales order data, distributor reports, or retail audit databases. Each unique store location receiving at least one unit of product within the measurement period is counted as one door Simple integer count of unique store identifiers (Store ID) with at least one confirmed sale or active SKU listing within the defined time period (monthly, quarterly, annually). Cross-validated against POS data and distributor shipment records Threshold: Absolute count (integer ≥ 0); Benchmarks vary by brand stage — Early Stage: <500 doors, Growth Stage: 500–5,000 doors, Scale Stage: >5,000 doors; National mass-market threshold typically ≥20,000 doors. Proposed by: Claude.

`retail_presence` *(Binary)* A binary indicator denoting whether a brand or product has any active retail distribution presence (physical or online retail channel) within a specified market or region during the measurement period Determined by confirming at least one active SKU listed and available for purchase in a retail channel within the defined geographic market and time window. Sourced from distributor records, retailer portals, or third-party retail audit data Binary classification: 1 = Product confirmed active in at least one retail door or online retail listing; 0 = No confirmed retail activity. Logistic regression or probit models used when modeling retail presence as a dependent variable Threshold: Binary: 0 (absent) or 1 (present); Presence confirmed when retail_doors ≥ 1 and at least one unit sold within the measurement period. Proposed by: Claude and GPT.

`retailer_interest` *(Ordinal (0–4))* Degree of expressed interest from retailers in stocking, trialing, or discussing the product. Assign a score based on retailer inquiries, meetings, negotiations, or stated willingness to carry the product. Survey/interview-based ordinal scoring on a 5-point scale. Threshold: 0=none, 1=low, 2=moderate, 3=high, 4=very

high retailer interest. Proposed by: GPT.

`retention` *(Numeric)* Share of customers, employees, or units retained over a specified period, typically referring to customer retention. Calculate as retained customers divided by customers at risk of churning over the period, multiplied by 100 if expressed as a percentage. Cohort retention analysis. Threshold: 0 = no retention; 100 = full retention when measured in percent. Proposed by: GPT.

`retention_rate` *(Numeric)* Percentage of customers, users, or accounts that remain active over a defined period relative to the cohort at the start of that period. Calculate as ((number of customers retained at end of period) / (number of customers at start of period, excluding new acquisitions if cohort-based)) x 100. The period should be consistently defined, such as monthly, quarterly, or annually. Cohort analysis using customer/account-level panel data from CRM, subscription, or transaction systems. Threshold: 0-100%, where higher values indicate stronger retention; if converted to an ordinal_0_4 scale: 0=<20%, 1=20-39%, 2=40-59%, 3=60-79%, 4=>=80%. Proposed by: GPT.

`return_on_ad_spend` *(Numeric)* Revenue generated for each monetary unit spent on advertising. Calculate as attributable revenue divided by advertising spend over the same campaign or reporting period. Marketing attribution analysis using campaign-level revenue and ad platform expenditure data. Threshold: ROAS >=0; common interpretation bands: <1.0 unprofitable, 1.0-1.99 weak, 2.0-3.99 moderate, 4.0-5.99 strong, >=6.0 very strong. Proposed by: GPT.

`return_rate` *(Numeric)* The proportion of sold units or revenue that is returned by customers or retailers within a defined period, expressed as a percentage of total units sold or gross revenue, indicating product quality, customer satisfaction, or demand forecasting accuracy Calculated from returns management systems, ERP data, or retailer chargeback records. Returns are matched to original sales transactions within the measurement window. Distinguishes between consumer returns, retailer returns, and defective product recalls Return Rate (%) = (Total Units Returned / Total Units Sold) × 100; or Revenue-based: (Value of Returns / Gross Revenue) × 100. Analyzed using time-series decomposition or control charts (SPC) to identify anomalies; regression used to identify drivers Threshold: 0–100%; Acceptable: <2%, Elevated: 2–5%, High (requires investigation): >5%, Critical (operational risk): >10%; Industry benchmarks vary — CPG/Beverage typically <1%, Apparel 20–30% acceptable. Proposed by: Claude and GPT.

`runway` *(Numeric)* Estimated amount of time a firm can continue operating before exhausting available cash, given its current net cash burn rate. Calculate as cash and cash equivalents divided by average net monthly burn rate. Cash flow forecasting and liquidity analysis using treasury balances and operating cash outflow data. Threshold: Measured in months, >=0; common bands: <3 critical, 3-5 short, 6-11 moderate, 12-17 comfortable, >=18 strong. Proposed by: GPT.

`sales_channel` *(Ordinal (0–4))* An ordinal classification of the route-to-market through which products are sold to end consumers, ranked by channel maturity, margin profile, and strategic value from least developed (0) to most developed or highest-value channel (4) Channels are classified and ranked based on a combination of margin contribution, volume scale, brand control, and strategic alignment. Classification assigned per SKU-channel combination based on active sales records, distributor agreements, and channel audit data. Ordinal scale constructed via expert scoring and validated against revenue contribution analysis Ordinal encoding: 0 = No active channel / Direct-to-Consumer only (nascent); 1 = Local/Independent retail or farmers markets; 2 = Regional grocery or specialty retail chains; 3 = National retail chains or e-commerce platforms (Amazon, DTC website); 4 = Mass-market retail (Walmart, Target, Costco) or omnichannel presence. Analyzed using ordered logit/probit regression; Spearman rank correlation used for association analysis with revenue outcomes Threshold: Ordinal scale 0–4: 0 = No formal channel, 1 = Local/Indie (1–50 doors), 2 = Regional chain (51–500 doors), 3 = National specialty/e-commerce (501–5,000 doors), 4 = Mass-market/omnichannel (>5,000 doors or national mass retailer); upgrade threshold: sustained 3-month revenue performance at current tier before reclassification. Proposed by: Claude and GPT.

`sales_channels` *(Ordinal (0–4))* The number and diversity of distribution channels through which a company sells its products or services, reflecting market reach and revenue diversification Score based on presence and activity across distinct sales channels including direct-to-consumer, wholesale, online marketplace, retail partnerships, and B2B enterprise sales Additive scoring: 0 = no established channel; 1 = single channel only; 2 = two distinct channels; 3 = three distinct channels with moderate revenue split; 4 = four or more channels with diversified revenue streams Threshold: 0: No active sales channel; 1: One channel (e.g., own website only); 2: Two channels (e.g., website + Amazon); 3: Three channels (e.g., website + Amazon + wholesale); 4: Four or more channels with documented multi-channel revenue. Proposed by: Claude.

`sales_efficiency` *(Numeric)* Revenue generated relative to sales and marketing expenditure, indicating how efficiently commercial spending produces sales outcomes. Calculate as incremental or recurring revenue generated divided by sales and marketing expense over the same or appropriately lagged period. Ratio analysis combining income state-

ment expenditure data with sales performance metrics; often implemented as SaaS magic number or revenue-to-S&M ratio. Threshold: Typically >=0; common bands: <0.5 weak, 0.5-0.99 below target, 1.0-1.49 acceptable, 1.5-1.99 strong, >=2.0 very strong. Proposed by: GPT.

`sales_mix` *(Numeric)* The proportionate distribution of total sales across products, channels, customer segments, or geographic markets during a defined period. Calculate the share of each category in total sales using revenue or unit sales. Aggregate category sales over the observation window and divide each category by total sales to obtain mix percentages. Ratio measurement using accounting or transaction data; optionally compute concentration metrics such as top-category share or Herfindahl-Hirschman style dispersion index. Threshold: 0-1 as category share of total sales; if using percentage form, 0-100. For concentration interpretation: >0.60 top-category share indicates highly concentrated mix, 0.30-0.60 moderate, <0.30 diversified. Proposed by: GPT.

`scalability` *(Ordinal (0–4))* The capacity of the business model to grow revenue significantly without a proportional increase in operational costs, reflecting leverage and infrastructure readiness Assessed through gross margin trajectory, operational leverage indicators, technology infrastructure, supply chain flexibility, and evidence of unit economics improvement at scale Composite scoring: 0 = no scalable infrastructure; 1 = limited scalability with high variable costs; 2 = moderate scalability with some fixed-cost leverage; 3 = strong scalability with proven unit economics improvement; 4 = highly scalable model with demonstrated exponential revenue vs. linear cost growth Threshold: 0: Purely manual/bespoke model; 1: Gross margin <30% with no improvement trend; 2: Gross margin 30-50% with some automation; 3: Gross margin 50-70% with documented cost leverage; 4: Gross margin >70% or demonstrated 2x+ revenue growth with <1.5x cost growth. Proposed by: Claude and GPT.

`seasonality` *(Ordinal (0–4))* The degree to which sales, demand, or operational activity systematically fluctuate by season, month, quarter, or holiday cycle. Compare peak-period sales to average or trough-period sales over at least one full annual cycle. Score the intensity and predictability of recurring fluctuations. Time-series decomposition, seasonal index calculation, or peak-to-average ratio analysis, converted into a 0-4 severity scale. Threshold: 0=no meaningful seasonal pattern; 1=mild seasonality with peak/average ratio <1.2; 2=moderate with ratio 1.2-1.5; 3=strong with ratio 1.5-2.0; 4=extreme with ratio >2.0 or heavy dependence on a narrow seasonal window. Proposed by: GPT.

`sell_through` *(Numeric)* The rate at which available inventory is sold to end customers within a specified period. Compute as units sold during the period divided by units available for sale, often beginning inventory plus receipts, or divided by inventory received for that cohort. Inventory turnover-style ratio using POS and inventory records. Threshold: 0-1 in ratio terms or 0-100 as percentage. Common interpretation: <0.40 weak, 0.40-0.70 moderate, >0.70 strong sell-through for the period. Proposed by: GPT.

`social_following` *(Ordinal (0–4))* The aggregate size and engagement quality of a brand's audience across social media platforms, serving as a proxy for organic marketing reach and community strength Aggregate follower count across primary platforms (Instagram, TikTok, Facebook, YouTube, Twitter/X) weighted by engagement rate; engagement rate = (likes + comments + shares) / followers Scoring based on total combined following and average engagement rate: 0 = <1K followers; 1 = 1K-10K followers; 2 = 10K-100K followers; 3 = 100K-500K followers; 4 = 500K+ followers or <100K with >5% engagement rate Threshold: 0: <1,000 total followers across all platforms; 1: 1,000-9,999 followers; 2: 10,000-99,999 followers; 3: 100,000-499,999 followers; 4: 500,000+ followers or documented viral presence with high engagement. Proposed by: Claude.

`social_mission` *(Ordinal (0–4))* The degree to which a company integrates a measurable social, environmental, or community-oriented mission into its core business model and operations Evaluated through mission statement formalization, B-Corp or equivalent certification, percentage of revenue or profit allocated to social causes, third-party impact reporting, and supply chain ethics documentation Scoring: 0 = no social mission; 1 = stated mission but no measurable commitment; 2 = documented commitment with partial implementation; 3 = certified or verified social impact with regular reporting; 4 = social mission is core to product/service with third-party verified impact metrics Threshold: 0: No mention of social mission; 1: Marketing language only, no commitments; 2: Stated goals with some evidence (e.g., 1% for the Planet pledge); 3: B-Corp certified or equivalent third-party verification; 4: Social mission embedded in legal structure (PBC/B-Corp) with published annual impact report. Proposed by: Claude.

`social_proof` *(Ordinal (0–4))* The strength of observable third-party validation of a product or firm, such as reviews, testimonials, ratings, user counts, endorsements, or case studies. Evaluate the breadth, credibility, and visibility of external validation signals across relevant channels. Assign score based on volume and quality of proof points. Composite rubric using review counts, average ratings, customer logos, influencer or expert endorsements, and public usage statistics. Threshold: 0=no meaningful social proof; 1=minimal proof with few testimonials or low visibility; 2=moderate proof with some reviews or references; 3=strong proof with many credible endorsements and positive ratings; 4=very strong proof with extensive, high-credibility, highly visible validation. Proposed by: GPT.

`strategic_fit` *(Ordinal (0–4))* The degree of alignment between the initiative, product, or company and the or-

ganization's long-term goals, capabilities, market positioning, and resource base. Assess alignment across strategic dimensions such as mission, core competencies, target market, synergies, and expected value creation. Score using a structured rubric. Multi-criteria expert assessment or weighted scorecard based on managerial interviews, planning documents, and capability mapping. Threshold: 0=no strategic alignment; 1=weak fit with major conflicts; 2=partial fit with mixed alignment; 3=good fit with clear synergies; 4=excellent fit, directly reinforcing core strategy and capabilities. Proposed by: GPT.

`strategic_partnerships` *(Ordinal (0–4))* The extent and quality of formal external collaborations that provide distribution, technology, supply, co-branding, or market access advantages. Evaluate number, relevance, exclusivity, and performance contribution of partnerships. Score based on both presence and strategic value of alliances. Structured qualitative scoring using contract status, partner caliber, revenue contribution, operational dependency, and exclusivity terms. Threshold: 0=no strategic partnerships; 1=informal or low-value partnerships only; 2=some formal partnerships with moderate value; 3=multiple valuable partnerships with clear business impact; 4=high-impact, defensible, and possibly exclusive partnerships central to growth. Proposed by: GPT.

`subscription_model` *(Binary)* Indicates whether the company generates recurring revenue through a formal subscription or membership model, providing predictable cash flow and customer retention metrics Binary classification based on the existence of an active, revenue-generating subscription or membership program with recurring billing cycles (monthly, quarterly, or annual) Binary assignment: 1 = subscription/membership model exists and generates >10% of total revenue; 0 = no subscription model or subscription revenue <10% of total revenue Threshold: 1: Active subscription program contributing ≥10% of total revenue with documented MRR/ARR; 0: No subscription program or de minimis subscription revenue (<10% of total revenue). Proposed by: Claude and GPT.

`subscription_rate` *(Numeric)* The proportion of customers, transactions, or revenue attributable to subscription plans within a defined period. Calculate as subscription customers divided by total customers, subscription transactions divided by total transactions, or subscription revenue divided by total revenue, depending on study design. Ratio estimation from billing and customer account data. Threshold: 0-1 as ratio or 0-100 as percentage. Interpretation: <0.10 low subscription penetration, 0.10-0.40 moderate, >0.40 high. Proposed by: GPT.

`supply_chain` *(Ordinal (0–4))* Degree to which the venture has established reliable sourcing, production, logistics, and supplier relationships to deliver its product or service at expected scale and quality. Assess the maturity and robustness of procurement, manufacturing, inventory, and distribution arrangements using documentary evidence such as supplier contracts, fulfillment records, lead-time history, and operational contingency plans. Expert scoring rubric based on operational due diligence and supply chain readiness assessment. Threshold: 0=no identified suppliers or delivery pathway; 1=informal or untested supplier/distribution arrangements; 2=basic supply chain in place but with limited redundancy or scale validation; 3=reliable and tested supply chain with documented partners and manageable risks; 4=highly robust, scalable, diversified, and consistently performing supply chain. Proposed by: GPT.

`team` *(Ordinal (0–4))* Overall strength of the founding and executive team in terms of capability to execute the venture's strategy. Evaluate the team holistically using founder backgrounds, role coverage, domain expertise, execution history, leadership capability, and evidence of cohesion drawn from resumes, interviews, references, and prior outcomes. Composite expert rating from venture screening or due diligence scorecard. Threshold: 0=team absent or clearly inadequate; 1=major capability gaps and weak execution evidence; 2=moderately capable team with some relevant skills but notable gaps; 3=strong team with relevant expertise and mostly complete role coverage; 4=exceptional team with highly complementary skills, strong track record, and proven execution. Proposed by: GPT.

`team_composition` *(Ordinal (0–4))* Extent to which the team has a balanced and complementary mix of functional, technical, commercial, and managerial capabilities required by the venture. Map current team members to critical venture functions such as product, engineering, sales, operations, and finance, and assess whether key competencies are represented and aligned to current stage needs. Capability-mapping rubric with expert evaluation of role complementarity. Threshold: 0=severely unbalanced team with most critical functions missing; 1=only one or two core competencies represented; 2=several needed functions covered but important gaps remain; 3=well-composed team with most critical functions covered; 4=highly complementary team with full functional coverage and strong alignment to venture needs. Proposed by: GPT.

`team_quality` *(Ordinal (0–4))* Quality of the team measured by the relevance of experience, prior achievements, educational/professional credentials, and demonstrated execution ability. Review team members' prior startups, industry experience, leadership roles, technical credentials, and performance outcomes, supplemented by references and observable milestones achieved by the current venture. Expert quality scoring using founder and management due diligence criteria. Threshold: 0=very weak credentials and no relevant execution evidence; 1=limited relevant experience or weak record; 2=adequate experience with some relevant accomplishments; 3=strong relevant background

and credible execution record; 4=outstanding credentials and repeated high-level success in closely relevant domains. Proposed by: GPT.

`team_size` *(Numeric)* The total number of full-time equivalent (FTE) employees and co-founders actively working in the business, reflecting operational capacity and human capital investment Count of all full-time employees plus 0.5 weighting for part-time employees (working <30 hours/week); contractors on long-term engagements (>6 months) counted as 0.5 FTE Continuous numeric variable: FTE = (full-time headcount) + (0.5 × part-time headcount) + (0.5 × long-term contractors); reported as a single decimal number Threshold: Minimum value: 1 FTE (solo founder); No upper bound; Values reported to nearest 0.5 FTE; Outliers defined as >3 standard deviations from industry-stage mean. Proposed by: Claude and GPT.

`technology_risk` *(Ordinal (0–4))* Level of uncertainty that the underlying technology will fail to perform, scale, integrate, or meet regulatory/market requirements. Assess technical feasibility, development stage, validation results, dependency on scientific breakthroughs, engineering complexity, and reproducibility based on prototypes, tests, audits, and technical documentation. Risk scoring through technical due diligence using a structured risk rubric. Threshold: 0=no meaningful technology risk, proven and validated technology; 1=low risk with minor unresolved technical issues; 2=moderate risk with some important validation or scaling questions; 3=high risk with major technical uncertainty; 4=very high risk with unproven core technology or substantial feasibility concerns. Proposed by: GPT.

`timing` *(Ordinal (0–4))* Degree to which the venture's market entry aligns with favorable industry, customer, technological, and regulatory conditions. Evaluate market readiness using indicators such as customer adoption trends, enabling infrastructure, regulatory environment, competitor timing, and macroeconomic context relative to the venture's offering. Expert market-timing assessment based on industry analysis and venture positioning. Threshold: 0=very poor timing, market not ready or already missed; 1=weak timing with substantial adoption barriers; 2=neutral or uncertain timing; 3=good timing with favorable market conditions; 4=excellent timing with strong convergence of demand, infrastructure, and enabling conditions. Proposed by: GPT.

`traction` *(Ordinal (0–4))* Observed market validation and momentum demonstrated through customers, revenue, user growth, retention, partnerships, or other measurable adoption signals. Review operating metrics such as monthly revenue, active users, growth rates, pilot conversions, churn, repeat purchases, and signed contracts to determine the extent of real market pull. Milestone-based scoring using venture performance data and customer validation evidence. Threshold: 0=no traction or market validation; 1=limited early signals such as pilots or small user base; 2=moderate traction with initial repeatable adoption; 3=strong traction with meaningful growth or revenue; 4=exceptional traction with clear product-market fit and sustained expansion. Proposed by: GPT.

`trade_secrecy` *(Ordinal (0–4))* Extent to which the venture's proprietary knowledge is protected through secrecy practices rather than formal intellectual property registration. Assess existence and enforcement of confidentiality procedures, access controls, employee and partner NDAs, compartmentalization of sensitive knowledge, and operational policies for protecting trade secrets. Legal/organizational protection scoring based on IP management due diligence. Threshold: 0=no meaningful secrecy protection; 1=minimal informal secrecy measures; 2=basic documented confidentiality practices; 3=strong trade secret controls with regular enforcement; 4=very strong, systematic, and well-enforced trade secrecy regime integrated into operations. Proposed by: GPT.

`unit_economics` *(Ordinal (0–4))* Assessment of whether the business demonstrates attractive contribution economics at the unit level, such as positive contribution margin, efficient customer acquisition relative to lifetime value, and scalable payback dynamics. Score based on evidence from gross margin, contribution margin, CAC-to-LTV relationship, payback period, and cohort-level profitability. Assign higher values when unit-level profitability is already proven and stable across cohorts or channels. 0=no evidence of viable unit economics; 1=weak or highly negative unit economics; 2=mixed or improving unit economics with unclear scalability; 3=solid positive unit economics with acceptable payback; 4=strong and repeatable best-in-class unit economics. Threshold: 0: contribution economics absent or materially negative; 1: negative contribution margin or CAC payback >24 months; 2: contribution margin near break-even or improving, payback 12-24 months; 3: positive contribution margin with payback 6-12 months and LTV/CAC >2x; 4: strong positive contribution margin with payback <6 months and LTV/CAC >3x. Proposed by: GPT.

`use_of_funds` *(Ordinal (0–4))* Quality and clarity of the company's planned deployment of raised capital across growth, operations, hiring, product, and runway extension. Evaluate whether capital allocation is specific, milestone-linked, realistic relative to burn and scale objectives, and aligned with value creation. Higher scores reflect disciplined and credible deployment plans. 0=no clear use of proceeds; 1=vague spending categories without milestones; 2=moderately clear allocation with partial milestone linkage; 3=clear and realistic allocation tied to operating plan; 4=highly detailed, disciplined, and milestone-driven capital deployment plan. Threshold: 0: no budget or stated plan; 1: broad

categories only; 2: line items exist but weak linkage to outcomes; 3: budget tied to hiring/product/go-to-market milestones and runway targets; 4: detailed allocation with quantified milestones, contingency planning, and explicit ROI logic. Proposed by: GPT.

`valuation_gap` *(Numeric)* Difference between the company's asked or current valuation and a benchmark valuation implied by fundamentals, comparables, or investor underwriting. Compute as company valuation minus benchmark valuation, or as a percentage premium/discount relative to the benchmark. Ensure benchmark construction is consistent across firms. Use either absolute gap = asked valuation - benchmark valuation, or relative gap = (asked valuation - benchmark valuation) / benchmark valuation. Positive values indicate over-benchmark pricing; negative values indicate under-benchmark pricing. Threshold: Continuous value; absolute gap can take any monetary value, relative gap can be any percentage. Common interpretation: <0 undervalued, =0 fairly valued, >0 over benchmark. Proposed by: GPT.

`viral_traction` *(Ordinal (0–4))* A composite ordinal measure capturing the degree to which a product, campaign, or content organically propagates across social networks and digital platforms through user-driven sharing, engagement, and word-of-mouth amplification beyond paid distribution Construct a composite index from five sub-indicators: (1) share-to-impression ratio, (2) organic reach growth rate, (3) user-generated content volume, (4) referral traffic coefficient, and (5) social mention velocity; normalize each sub-indicator to a 0–1 scale using min-max normalization; compute weighted average score; map final score to ordinal scale using predefined percentile-based thresholds derived from baseline benchmark data Ordered logistic regression (proportional odds model) for inferential analysis; Spearman rank correlation for association testing; Kendall's tau for concordance measurement; principal component analysis (PCA) to validate composite index dimensionality; use Mann-Whitney U test for group comparisons given ordinal nature of the variable Threshold: 0 = No viral traction (composite score 0.00–0.19; organic reach growth <5%, share rate <0.5%); 1 = Minimal traction (score 0.20–0.39; organic reach growth 5–15%, share rate 0.5–1.5%); 2 = Moderate traction (score 0.40–0.59; organic reach growth 16–35%, share rate 1.6–3.0%); 3 = Strong traction (score 0.60–0.79; organic reach growth 36–75%, share rate 3.1–6.0%); 4 = Viral breakout (score 0.80–1.00; organic reach growth >75%, share rate >6.0%). Proposed by: Claude.

`vision` *(Ordinal (0–4))* Strength, ambition, and coherence of the founder's long-term strategic narrative regarding the problem, market transformation, and company trajectory. Assess whether the vision is clear, differentiated, plausible, and connected to a large market opportunity and concrete execution path. Higher scores indicate compelling and internally consistent strategic direction. 0=no coherent vision; 1=generic or weak vision; 2=moderately clear but limited differentiation; 3=clear, ambitious, and credible vision; 4=exceptionally compelling, differentiated, and market-shaping vision. Threshold: 0: absent or inconsistent narrative; 1: broad aspiration without strategic logic; 2: understandable direction but limited distinctiveness; 3: strong articulation of future state, customer value, and scaling path; 4: highly persuasive, differentiated long-term thesis with strong founder-market fit and strategic coherence. Proposed by: GPT.

`waitlist` *(Numeric)* Number of prospective users, customers, or accounts that have expressed interest and are queued for access, onboarding, or product availability. Measure as the count of unique qualified sign-ups or prospective customers on the waitlist at a specified date. Deduplicate entries and define qualification criteria consistently. Record the total unique waitlist count. Where useful, supplement with growth rate, conversion rate from waitlist to active users, and share of qualified leads. Threshold: Continuous count variable; minimum 0. Example interpretation only: 0=no waitlist, low=1-999, medium=1,000-9,999, high=10,000+ if the study later bins values. Proposed by: GPT.

`wholesale_mix` *(Numeric)* Share of total sales derived from wholesale channels relative to all company revenue. Compute wholesale revenue divided by total revenue over a defined reporting period, ensuring revenue recognition is consistent across channels. Wholesale mix = wholesale revenue / total revenue, expressed as a percentage or decimal share. Threshold: Continuous proportion from 0 to 1, or 0% to 100%. 0=no wholesale exposure; 1 or 100%=all revenue from wholesale. Proposed by: GPT.

C.3 Reproducibility Matching Prompt

Section 6.9 and Appendix B.1 report how far an independently generated variable list returns the paper's variables. Counting a return by name is mechanical; counting one by meaning is not, and this is the prompt that does it. Claude is asked, for every directed pair of variable-generation runs in Table B.8, which variables describe the same underlying concept. Each run serves as List A once against every other run, so that a variable's presence or absence is checked from both lists' perspective, and {list_a} and {list_b} are filled with the two runs' variable names and one-line definitions.

Prompt

```
MATCH_PROMPT = You are comparing two lists of variables that were independently generated
by a large language model to describe early-stage startup investment pitches.  Each
variable has a name and a one-sentence definition.  Some variables in List A describe
the exact same underlying concept as a variable in List B, even though the two lists
were generated separately and may use different names or slightly different wording
for the same idea.  Other variables in List A have no real counterpart in List B at
all.

Your task:  for each variable in List A, decide whether there is a variable in List
B that describes the SAME underlying concept (not just a related or overlapping topic
-- the same concept).  Be conservative:  two variables about ‘‘risk’’ in general are
not a match just because they share the word ‘‘risk’’ -- they must describe the same
specific thing (e.g., both about execution risk specifically, not one about execution
risk and one about regulatory risk).

List A:
{list_a}

List B:
{list_b}

Return ONLY a JSON array, one object per variable in List A, in this exact format,
and nothing else before or after the JSON:
[
{"variable_a":  "<name from List A>", "matched_variable_b":  "<name from List B, or
null if no match>", "confidence":  "<high|medium|low>"}
]
```

C.4 Post-show outcome collection

Every post-show outcome in this paper comes from the open-web collection described here: completion of the handshake, operating status, survival at one, two and three years, the year of closure, acquisition with its year and acquirer, whether the venture raised after airing, whether that round was institutional, and the amount raised.

Acquisition and follow-on funding are also available from Crunchbase, and we collect them there as well. The two records are kept apart and never merged, since they carry different selection; the open-web collection covers all 642 ventures, a Crunchbase profile exists for 388.

C.4.1 The open-web collection

Two passes ran in September 2026, both on `claude-opus-5`. Each split the 642 ventures into 65 batches of at most ten and gave every batch to a fresh agent with no memory of any other batch, no access to our data, and no knowledge of the other pass. An agent received the venture's name, season, product, founder state, company website and on-air deal terms; searched the open web with search and page retrieval only; recorded the address of every page it opened; and coded each field from those pages alone, never from prior knowledge. Crunchbase and the sites that republish it were out of bounds, so the collection is independent of the Crunchbase record. Agents opened 6.9 sources per venture in the first pass and 7.3 in the second. The instruction is reproduced below.

Within a pass, a value was recorded wherever the agent could establish it, with a separate flag for whether two independent sources supported it. Independence is defined over five families: sites that recap the show, which count as one source however many are read; the venture itself; the investor; databases and public filings; and independent news outlets, each counting separately. A recap repeating another recap is not a second source.

Across passes, a value is retained only where both returned the same reading. Where they disagreed, or where only one returned anything, the field is left missing. Survival is additionally treated as established where the venture was recorded as operating at the collection date and that date lay beyond the horizon. Table C.15 gives the basis of every value.

Evidence depth is uneven. Of the 593 retained operating statuses, 520 rest on two independent sources in both passes and 23 on a single source in both; of the 313 retained completion outcomes, 67 rest on two independent sources in both passes and 147 on a single source in both. Completion is the thinnest-sourced field and the one the passes disagree on most often.

C.4.2 Sources opened

Source type	Citations	Share	Domains
Shark Tank follow-up sites	763	21.9%	16
General-interest and entertainment sites	670	19.2%	38
Business and trade press	260	7.5%	29
Local news and founder interviews	196	5.6%	123
Company databases and public registries	177	5.1%	17
Reference works	61	1.8%	1
Press releases	44	1.3%	4
Crowdfunding and marketplace listings	27	0.8%	6
Social and professional networks	9	0.3%	4
Other individual sites	797	22.9%	604
Ventures' own websites	477	13.7%	475

Table C.15: Details of open-web collection

Field	Basis	Pitches	Retained
Completion	Both passes agree	313	313
	The two passes disagree	21	0
	No *handshake*, so nothing to resolve	213	0
	Neither pass established a value	27	0
	Pass 1 only	44	0
	Pass 2 only	24	0
Operating status	Both passes agree	593	593
	The two passes disagree	25	0
	Neither pass established a value	4	0
	Pass 1 only	11	0
	Pass 2 only	9	0
Operating after 1 year	Both passes agree	615	615
	Implied by operating status at the collection date	4	4
	The two passes disagree	5	0
	Pass 1 only	13	0
	Pass 2 only	5	0
Operating after 2 years	Both passes agree	533	533
	Implied by operating status at the collection date	7	7
	The two passes disagree	2	0
	Neither pass established a value	7	0
	Horizon not yet elapsed at the collection date	70	0
	Pass 1 only	15	0
	Pass 2 only	8	0
Operating after 3 years	Both passes agree	490	490
	Implied by operating status at the collection date	8	8
	The two passes disagree	4	0
	Neither pass established a value	8	0
	Horizon not yet elapsed at the collection date	108	0
	Pass 1 only	11	0
	Pass 2 only	13	0
Acquired	Both passes agree	587	587
	The two passes disagree	4	0
	Neither pass established a value	6	0
	Pass 1 only	25	0
	Pass 2 only	20	0
New investor round	Both passes agree	371	371
	The two passes disagree	15	0
	Neither pass established a value	58	0
	Pass 1 only	116	0
	Pass 2 only	82	0
Year of closure	Both passes agree	53	53
	The two passes disagree	6	0
	Neither pass established a value	548	0
	Pass 1 only	21	0
	Pass 2 only	14	0

C.4.3 Collection prompt

The instruction below was given to every agent in both passes, reproduced verbatim; {BATCH} and {OUT} were replaced with that batch's input and output paths.

```
You are collecting post-show outcome data on US Shark Tank ventures from the open web.
You have NO prior data on these companies.  Do not code anything from memory or prior
knowledge.
Work only from pages you open in this session.

COLLECTION DATE: treat today's date as the collection date.  Judge "as of today"
against it.

INPUT: read the CSV at {BATCH}.  Columns:  uid, company, product_name, season,
airing_year,
airing_date, founder_state, company_website, onair_deal.  'onair_deal' says whether the
founders
and an investor shook hands on air.  It does NOT say whether the money ever arrived.

================================================================================
FIELDS TO COLLECT
================================================================================
For EACH venture collect all of the following.

deal_closed yes | no | unknown
Did the on-air handshake actually result in the investment being made?
"no" = the agreed deal was abandoned, renegotiated away, or never funded.
Leave unknown where onair_deal is "no".

operating_status active | closed | acquired | unknown Status as of the collection date.

alive_year1 yes | no | unknown Operating 1 year after the airing_date?
alive_year2 yes | no | unknown Operating 2 years after?
alive_year3 yes | no | unknown Operating 3 years after?

closure_year YYYY | unknown Year it ceased trading, if it did.

acquired yes | no | unknown Was the venture ever acquired or did it exit?
acquisition_year YYYY | unknown
acquirer name | unknown

raised_capital_after_airing yes | no | unknown
Did the venture take in ANY outside capital after the episode aired?
This INCLUDES the investor's own money if the on-air deal closed.
It includes angel money, crowdfunding that funded, debt, grants,
accelerator money and institutional rounds.

institutional_round yes | no | unknown
Did the venture raise a named, institutional-style round after airing
(seed, Series A or later, or a venture fund taking a stake)?
A closed Shark Tank deal on its own does NOT count here.
This is a STRICTER subset of raised_capital_after_airing.

total_raised_usd number | unknown Total outside capital raised after airing, in USD.

================================================================================
SOURCES AND THE TWO-SOURCE RULE
================================================================================
Source families:
```

```
A recap sites - sharktankblog, sharktanktales, sharktankrecap, looper, mashed and
similar.
ALL of family A counts as ONE source no matter how many you read.
B the company itself - its site, shop, blog, socials, press releases
C the investor - portfolio page, their own posts
D databases and filings - SEC or state filings, Wefunder, StartEngine, Kickstarter,
patent records, business registries
E independent news - each outlet counts separately

Two sources are independent if they are in two different families, or two different
outlets
within family E. A recap repeating another recap is NOT a second source.

DO NOT USE CRUNCHBASE for this task.  Do not open crunchbase.com and do not code
anything from
it.  Sites that merely republish Crunchbase data (PitchBook, CB Insights, Craft,
ZoomInfo,
Owler, RocketReach) are also excluded.  This collection has to stand on its own.

FOR EVERY FIELD, ALWAYS TRY TO FIND TWO INDEPENDENT SOURCES.
- If you find two, record both and set the field's *_2src column to "yes".
- If you find only one, STILL RECORD THE VALUE. Put the source in src1, leave src2
empty,
and set *_2src to "no".  Do not downgrade the value to unknown for want of a second
source.
- Only use "unknown" when you could not establish the value at all.
The value and the strength of its evidence are two separate things.  Record both.

SEARCH WITHOUT LIMITS. Use as many searches and page fetches as the task needs.
Nothing is
rationed.  Try the company website, "<company> shark tank update", "<company> shark
tank deal
closed", "<company> out of business", "<company> acquired", "<company> raises",
"<company>
funding", the founder's name, and local and trade press.  A dead website or a shop that
no
longer loads IS usable evidence, from family B.

================================================================================
OUTPUT
================================================================================
Write a CSV to {OUT} with exactly these columns, one row per venture, in the input
order:

uid, company,
deal_closed, deal_closed_2src, deal_closed_src1, deal_closed_src2,
operating_status, operating_status_2src, operating_status_src1, operating_status_src2,
alive_year1, alive_year2, alive_year3, survival_2src, survival_src1, survival_src2,
closure_year, closure_year_2src, closure_year_src1, closure_year_src2,
acquired, acquisition_year, acquirer, acquired_2src, acquired_src1, acquired_src2,
raised_capital_after_airing, raised_capital_2src, raised_capital_src1,
raised_capital_src2,
institutional_round, institutional_2src, institutional_src1, institutional_src2,
total_raised_usd, total_raised_src1,
all_sources, notes

Rules for the output:
- Put the actual URL in every *_src column, never a description.
- all_sources = every URL you opened, separated by " | ".  Include pages that turned out
```

```
useless; the record has to show what was looked at.
- *_2src is "yes" or "no" and refers to whether TWO independent sources were found.
- notes = anything a reader would need to interpret the row:  contradictions between
sources,
a deal renegotiated rather than abandoned, an acquisition of assets rather than the
company.
- Write the file even if most values are unknown.

REPLY FORMAT. Your entire final message must be ONE line:
"batch done:  N rows, deal_closed established X, raised_capital established Y"
Nothing else.  No per-venture summary, no findings, no caveats.  Everything you want to
say
belongs in the ‘notes‘ column, not in your reply.
```