Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

When Expert Screening and Crowd Validation Disagree: Signal Discordance and Venture Quality in Life-Science Equity Crowdfunding

Laura Sáez-Ortuño
University of Barcelona
Marti Sagarra
martisagarra@ub.edu
University of Barcelona
Santiago Forgas-Coll
University of Barcelona
Valeria Isgrò
University of Reggio Calabria
Massimiliano Ferrara
University of Reggio Calabria

Research Article

Keywords: Equity crowdfunding, Expert screening, Signal discordance, Crowd dynamics, Venture quality, Entrepreneurial finance

Posted Date: April 29th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9509733/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

When Expert Screening and Crowd Validation Disagree: Signal Discordance and Venture Quality in Life-Science Equity Crowdfunding

Abstract

Specialized equity-crowdfunding platforms generate three non-equivalent signals about each venture: expert screening before launch, visible early traction during the campaign, and venture outcomes observed years later. These signals are often studied separately, even though platforms and investors must interpret them jointly, especially when they disagree. Using ninety-nine rounds closed on Capital Cell, a life-science equity-crowdfunding platform, between 2016 and 2023, we study this problem as one of *signal discordance*. We build block-level measures that compare expert evaluations, crowd traction, and lead-investor dominance, and we relate their agreement or disagreement to post-campaign venture quality, measured through venture status and valuation multiples. To analyse a small, noisy, and block-structured dataset under epistemic uncertainty, we use a soft quantum kernel whose softness is calibrated from the dispersion of expert scores. The method is used here as a tool for mixed-signal settings rather than as a claim of general quantum superiority, and we benchmark it against classical alternatives and a hardware implementation on the Quantum Blue QPU at the Barcelona Supercomputing Center. The results show that expert screening and crowd traction are distinct informational objects, that their disagreement is especially informative about long-run quality, and that specialized platforms add value when expert signals remain informative where early market validation does not.

Plain English Summary. Fast fundraising does not necessarily mean high venture quality. Some companies attract strong early momentum on a crowdfunding platform and later underperform, while others raise more slowly but create more durable value. Using ninety-nine Capital Cell rounds in life sciences, this paper asks what happens when expert screening before launch and crowd behaviour during the campaign point in different directions. We find that expert assessments contain useful information about long-run outcomes and that the most revealing cases are precisely those in which expert and crowd signals disagree. For investors and platform designers, the implication is straightforward: early traction is informative, but it is not a substitute for specialised due diligence.

Keywords: Equity crowdfunding; Expert screening; Signal discordance; Crowd dynamics; Venture quality; Entrepreneurial finance

JEL Classification: G24; G14; L26; D82

1 Introduction

Equity crowdfunding has developed from a marginal experiment in entrepreneurial finance into an established channel for ventures that are too young, too opaque, or too risky to raise capital from banks

or conventional early-stage investors (Ahlers et al., 2015; Bruton et al., 2015; Cumming et al., 2021). This role is especially relevant for SMEs facing financial constraints and agency frictions that make alternative financing channels particularly attractive (Fasano et al., 2025). Its central analytical interest lies in the severity of the information asymmetry under which investment decisions are made: a broad and heterogeneous crowd, much of it unsophisticated, must evaluate early-stage ventures on the basis of limited and often ambiguous disclosure. A large body of work has examined the signals that reduce this asymmetry, including retained equity, valuation, founder characteristics, third-party endorsements, communication, and funding dynamics (Ahlers et al., 2015; Courtney et al., 2017; Piva and Rossi-Lamastra, 2018; Ralcheva and Roosenboom, 2020; Vismara, 2016).

A parallel line of research has examined the role of the platform itself as a screening device. Specialised platforms do not operate as neutral marketplaces but as active gatekeepers that screen, certify, and sequence ventures before a campaign is launched (Cumming et al., 2019; Kleinert et al., 2022). On sector-specialised platforms this screening function becomes particularly consequential, because the underlying technology, intellectual property, and scientific risk are often beyond the evaluative capacity of dispersed retail investors. Capital Cell, the platform from which our data are drawn, is one such case. It operates in biotech, medtech, and digital health, and it subjects every venture to a formal pre-campaign review by the BioExpert Network, followed by internal and financial due diligence and a confirmed lead-investor commitment before launch (Capital Cell, 2025a,b).

The combination of these two features, a specialised platform that produces an expert signal before the campaign opens and a visible pattern of crowd and lead-investor behaviour during the campaign itself, generates a research setting that is rarely accessible to entrepreneurial-finance researchers. Most empirical work in the field has been forced to treat ventures as observable only through a single lens, either the information disclosed at launch or the dynamics of the campaign or the outcomes recorded afterwards. In the Capital Cell data, all three coexist at the round level for ninety-nine campaigns closed between 2016 and 2023, together with post-campaign venture status and valuation information as of the end of 2025. This gives us a rare opportunity to pose a question that the existing literature has largely had to sidestep: when expert screening and crowd validation disagree about a venture, which source of information is more informative about what happens afterwards?

The question matters for reasons that speak directly to the entrepreneurial-finance agenda of this journal. Platforms and their screening technologies are increasingly understood as active intermediaries whose selection decisions shape the composition of the crowdfunding market (Kleinert et al., 2022; Correia et al., 2024). The distinction between early campaign success and longer-run venture outcomes has been shown, in settings where both are observable, to be substantively large (Signori and Vismara, 2018; Hornuf et al., 2018; Walthoff-Borm et al., 2018; Coakley et al., 2022; Valenza et al., 2023). Recent evidence has also begun to document that visible investor actions on platforms, including those of platform-affiliated teams themselves, can produce signals that mislead rather than inform (Mataigne et al., 2025). Taken together, these lines of work point to a single underlying problem: the information environment of equity crowdfunding is structured into distinct blocks that are only partially correlated, and the theoretical and empirical treatment of the field has not yet caught up with this multiplicity.

This paper takes the view that the information environment of a specialised equity-crowdfunding platform should be analysed as a system of partially overlapping signal blocks, and that the most substantively informative empirical cases are those in which the blocks disagree. We construct indices of

discordance between the expert block, the crowd-dynamics block, and the lead-investor block, and we ask how discordance maps onto long-run venture quality measured through status and valuation multiples. The ventures on which expert opinion and early traction diverge are not anomalies to be averaged away; they are the ventures that reveal whether specialised due diligence captures durable value that market behaviour alone cannot see.

Handling this question with the ninety-nine-round sample available to us requires an empirical tool with specific properties. It should accommodate small samples, preserve the ordinal nature of the expert signal, and, crucially, be able to represent the epistemic uncertainty of an expert panel whose inter-rater dispersion is informative rather than nuisance. The tool we adopt is a kernel method with a soft-quantum construction in which observations are represented not as precise feature vectors but as fuzzy states whose degree of sharpness is calibrated from the dispersion of the expert scores themselves. The construction, which we call a *soft quantum kernel*, interpolates continuously between a classical Gaussian kernel and a pure-state quantum kernel and is designed specifically for the small, structured, and noisy regime that characterises the Capital Cell data. Because the kernel is fully implementable in classical simulation, we can treat it as a conventional learning algorithm in the empirical analysis; because it admits a hardware analogue, we can also verify that its predictions survive execution on a physical quantum device. We conduct this verification on the Quantum Blue QPU at the Barcelona Supercomputing Center as a robustness check rather than as a methodological claim of quantum advantage.

The paper proceeds as follows. Section 2 positions the argument within the entrepreneurial-finance literature on screening, campaign dynamics, and post-campaign outcomes, and states the research questions. Section 3 describes the Capital Cell setting, the data, and the signal-discordance construction. Section 4 presents the soft quantum kernel verbally and with a single defining equation; the formal development is placed in Appendix A. Section 5 sets out the empirical strategy and the hardware validation. Section 6 discusses implications for theory, practice, and platform design, and concludes.

2 Screening, Crowd Dynamics, and the Problem of Conflicting Signals

2.1 Specialised platforms and the production of an expert signal

The theoretical starting point of our argument is Spence (1973)'s observation that, in markets with severe information asymmetry, a signal becomes valuable only when it is observable, credible, and costly to imitate. In entrepreneurial finance this logic is well established (Connelly et al., 2011), and in equity crowdfunding it has been used to interpret the predictive role of retained equity, founder human capital, third-party endorsements, and external validation (Ahlers et al., 2015; Vismara, 2016; Piva and Rossi-Lamastra, 2018; Ralcheva and Roosenboom, 2020). A more recent strand has shifted attention away from signals emitted by the venture itself and towards signals produced by the platform through its screening process. Specialised platforms do not admit all applicants; they select, curate, and certify (Cumming et al., 2019; Kleinert et al., 2022). When this selection is rooted in an expert panel with domain knowledge, the platform effectively produces a pre-campaign signal that was previously internal to venture capitalists and is now made available, at least indirectly, to the crowd.

On Capital Cell this production is institutionalised in the BioExpert Network, whose reviewers evaluate each candidate along the dimensions of team, science, innovation, and business model and deliver a composite overall grade. The overall grade is a bounded continuous score, typically in the range from

six to nine, and its dispersion across ventures is moderate: for the ninety-six non-missing observations in our sample the mean is 7.47 and the standard deviation is 0.53, with values ranging from 6.0 to 9.2. The compression of the score range is itself meaningful. The panel effectively filters out ventures that would be rated below a minimum threshold before they reach the campaign stage, so that the scores observed in the data are conditional on a first-stage acceptance decision that is already selective. This conditional character is a feature of the data, not a defect, and it is one of the reasons the expert signal cannot be analysed as if it were a noise-free measurement.

2.2 Campaign dynamics as a second signal layer

Once a campaign opens, a distinct and quite different informational process begins. Hornuf and Schwienbacher (2018) and Vismara (2018) have shown that early investor behaviour, including the presence of lead or early-bird investors, generates observable dynamics that shape subsequent participation, and Block et al. (2018) extend this logic to the effects of campaign updates and communication. Pommet et al. (2024) document that the commitment of lead investors is associated with both fundraising success and post-campaign liquidity, and Mataigne et al. (2025) show that visible actions by platform-team members can themselves generate false signals associated with weaker venture outcomes.

The Capital Cell data make this temporal process unusually transparent. By day three of the campaign, including the lead commitment, the average round has already reached 48.7% of its target and the median has reached 43.2%; by day seven the corresponding figures rise to 59.9% and 52.8%. The lead investor alone accounts, on average, for 40.9% of the total round, with a median share of 43.6%. What these numbers reveal is that early traction on Capital Cell is rapid and heavily shaped by a small number of committed actors whose decisions precede and anchor crowd participation. This is an important substantive observation because it means that early traction cannot be read as an unmediated measure of crowd enthusiasm. It is a composite signal in which lead-investor judgment and subsequent herding behaviour are intertwined.

2.3 Post-campaign outcomes as a third layer

A significant body of evidence now documents that the post-offering lives of crowdfunded firms are highly heterogeneous (Signori and Vismara, 2018; Hornuf et al., 2018; Walthoff-Borm et al., 2018; Butticè et al., 2020; Coakley et al., 2022; Valenza et al., 2023). Fundraising success and venture success, in the sense of survival, follow-on investment, or value creation, are related but far from identical. In our sample this pattern is sharp. Of the ninety-nine rounds in the general sheet, seventy-eight are classified as in business at the end of 2025, fifteen as written off, three as zombies, two as exits, and one as merged. The valuation multiple ranges from zero to just above twelve, with a median of exactly one and substantial right skewness. Twenty rounds have a valuation multiple of zero, reflecting either write-offs or ventures that have not raised at higher valuations since their campaign. This distribution makes linear measures of association such as Pearson correlation a very demanding summary of the expert–outcome relationship and points directly to the importance of rank-based and ordinal outcome transformations, a theme we return to in the empirical section.

2.4 From three signal blocks to a discordance framework

The three layers we have described correspond to three distinct informational moments in the life of a crowdfunding round. The expert block is generated before the campaign by a structured panel with domain expertise but limited incentive alignment. The crowd-dynamics block is generated during the campaign by a mix of professional leads, semi-professional early investors, and retail followers whose actions are observable in real time but whose motivations are difficult to recover. The outcomes block is generated over a horizon of several years by the venture's own trajectory and by subsequent funding markets. The standard move in the literature has been to treat these blocks as complementary inputs to a single predictive model of campaign success or, in more recent work, of post-campaign survival. This framing has been productive, but it obscures a feature of the data that is, in our view, more theoretically revealing: the blocks can and do disagree, and their disagreement is itself informative.

Consider two stylised cases that are both represented in the Capital Cell sample. A venture may receive a moderate expert score yet attract very rapid early traction, reaching a large share of its target within days and closing quickly. Read through the lens of the campaign-dynamics literature, this is a successful campaign. Read through the lens of expert screening, it is a borderline case. Over the longer horizon the two readings may converge, with the crowd having correctly identified value the experts understated, or they may diverge further, with the venture turning out to be a write-off despite its visible early momentum. A second venture may display the opposite pattern: a high expert score, a slow campaign, and a strong post-campaign trajectory measured by valuation growth or follow-on funding. These cases are not noise around a single underlying signal; they are the observations on which the marginal informational content of each block is most clearly identified.

We build this idea into the empirical design by constructing block-level indices of discordance, which compare the rank of each venture within the expert block to its rank within the crowd-dynamics block and within the lead-dominance block. The details of the construction are discussed in Section 3.4. The point we want to emphasise here is conceptual: discordance is the variable of interest because it operationalises the classical problem of signal conflict in a setting where three distinct production processes coexist for the same venture. Figure 1 summarises the framework.

2.5 Research questions

We formulate the empirical problem in terms of four research questions that stay close to the entrepreneurial-finance tradition of the journal and that are expressed as comparative statements about the blocks rather than as tests of an abstract predictive model.

> **RQ1.** To what extent do pre-campaign expert evaluations carry ordinal information about long-run venture quality, measured through post-campaign status and valuation multiples, once early traction is controlled for?
>
> **RQ2.** Is early campaign traction a distinct informational object that reflects lead-investor behaviour and crowd dynamics more than it reflects expert judgment?
>
> **RQ3.** Do the expert and crowd blocks contain complementary rather than redundant information about long-run quality, so that the joint model improves over either block alone precisely in the subsample where the two blocks disagree?

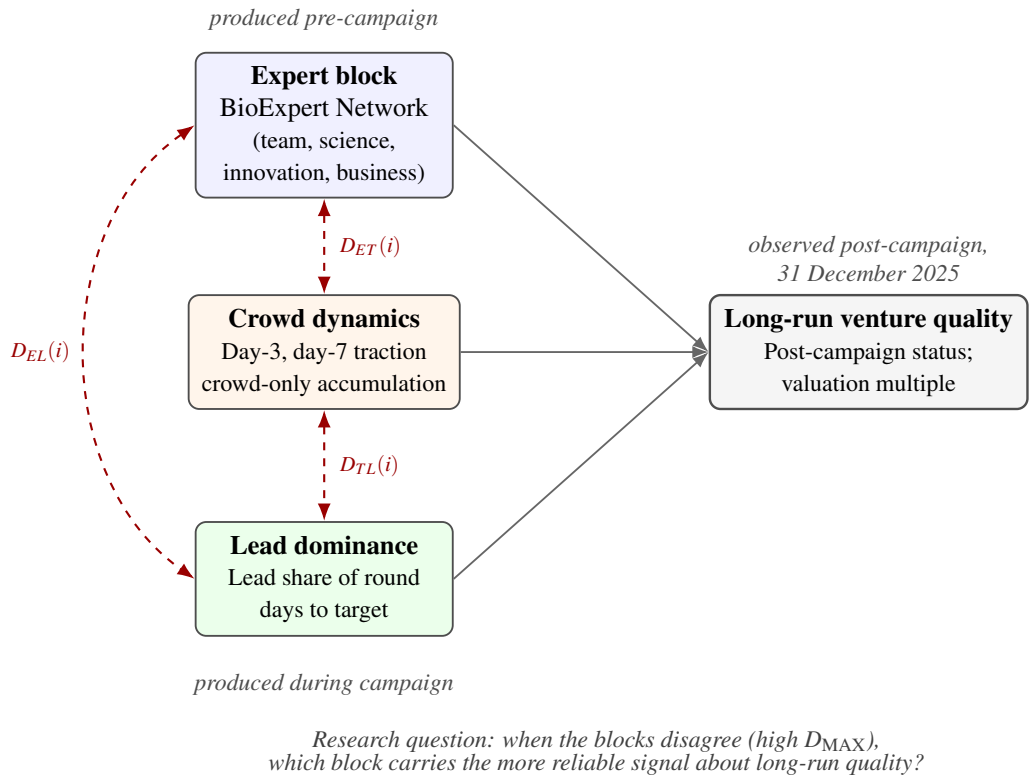


Figure 1: Signal-discordance framework. Three distinct signal blocks are produced at different moments in the life of a Capital Cell round: the expert block before the campaign, crowd dynamics and lead dominance during the campaign, and long-run venture quality several years later. The dashed arrows denote pairwise discordance indices $D_{ab}(i)$ that measure, for each venture i, the extent to which two blocks disagree in their relative ranking of that venture.

> **RQ4.** Does expert screening retain predictive power for pairwise rankings of venture quality, so that it can correctly identify which of two ventures will exhibit the higher long-run multiple even when standard regressions on the raw multiple show no linear association?

These questions share a common logic. They ask whether expert screening produces a signal that survives in regimes where standard predictive models, driven mainly by crowd behaviour, would not expect to find one. Framed in this way, the study sits in the intersection of the entrepreneurial-finance literature on signals and information asymmetry and the platform-design literature on curated marketplaces, and it contributes to the debate on when and why specialised intermediaries add value.

3 Data, Measurement, and the Construction of Signal Discordance

3.1 Institutional setting and data

Capital Cell is a Spanish-headquartered equity-crowdfunding platform specialising in the life sciences. Its evaluation pipeline begins with the BioExpert Network, proceeds through internal screening and financial and intellectual-property due diligence, and is completed only when a lead investor has formally committed to the round. We analyse the full universe of rounds closed between 2016 and 2023 for which a complete post-campaign outcome can be observed as of 31 December 2025, yielding ninety-nine round-level observations in the general sheet. A companion speed sheet provides finer-grained

Table 1: Descriptive statistics for the Capital Cell sample.

Variable	N	Mean	Median	SD	Min	Max
General sheet ($N = 99$ round-level observations)						
Overall Grade	96	7.47	7.50	0.53	6.00	9.20
Team score	99	7.99	7.90	0.60	6.80	9.50
Science score	99	7.95	7.90	0.61	6.60	9.50
Innovation score	93	7.66	7.70	0.57	6.30	9.30
Business score	99	7.60	7.60	0.61	6.40	9.10
Initial valuation (€M)	99	13.20	5.54	25.61	0.50	200.00
Website target (€K)	99	778.8	640.0	480.7	100.0	3000.0
Amount raised (€K)	99	532.8	391.1	446.0	40.7	2000.0
Duration (days)	99	62.8	58.0	41.2	1	156
Number of investors	99	244.7	190.0	179.5	12	801
Valuation multiple	99	1.472	1.000	1.685	0.000	12.055
Lead share of round	99	0.409	0.434	0.246	0.000	0.855
Speed sheet ($N = 102$ campaign-level observations)						
% target day 3 (incl. lead)	102	48.70	43.19	28.59	2.38	139.90
% target day 7 (incl. lead)	102	59.86	52.81	34.59	4.18	175.44
% raised by lead	102	40.86	43.61	25.25	0.10	85.46
Days to target / max duration	102	51.87	43.00	42.77	1	170

Notes. Expert scores are on the BioExpert Network 0–10 scale. Currency figures are in euros.

campaign-dynamics variables for one hundred and two campaign-level observations, including early percentage of target reached and the lead share of the final round. Sectoral concentration is high: biotech accounts for forty-three rounds and medtech for forty-two, with thirteen digital-health rounds and one industrial round. Geographically, ninety-seven of the ninety-nine rounds are domiciled in Spain. The cohort grows from five rounds in 2016 to nineteen in 2023.

Round-level descriptive statistics are reported in Table 1 for the main variables used in the analysis.

3.2 Dependent variables: traction and long-run quality

The argument developed in Section 2 requires that success not be collapsed into a single variable. We therefore work with two families of dependent variables. The early-traction family uses continuous measures drawn from the speed sheet, specifically the percentage of target reached by day seven including the lead and the ratio of days to target over maximum allowed duration. These variables measure how rapidly the round fills and are sensitive to early investor behaviour. The long-run quality family uses the valuation multiple and a harmonised binary indicator distinguishing ventures that have been written off from those that remain in business or have experienced a positive exit. Because the valuation multiple is highly right-skewed and has a mass at zero, we additionally analyse its log transformation on the subsample of ventures with strictly positive multiples, and we adopt a pairwise ranking formulation in which the outcome is whether a model correctly identifies the venture with the higher multiple in a randomly sampled pair. The pairwise construction is particularly appropriate for small noisy datasets because it is invariant to monotone transformations of the outcome and exploits information that is lost when one conditions on the raw metric value alone.

Table 2: Bivariate correlations between expert scores and long-run outcome measures.

	Raw VM		Log VM ($n = 79$)		High-growth
Variable	Pearson	Spearman	Pearson	Spearman	point-bis.
Overall Grade	0.016	0.296**	0.027	0.290*	0.279
Innovation	0.031	0.284**	0.134	0.335**	0.258
Business	0.095	0.191	0.168	0.328**	0.285
Combined All Scores	0.050	0.230*	0.105	0.305**	—
Team	0.055	0.110	0.067	0.169	0.196
Science	−0.026	0.100	0.018	0.160	0.160
Number of Reviews	−0.179	−0.047	−0.407**	−0.341**	—

Notes. * $p < 0.05$; ** $p < 0.01$. Raw VM: $n = 99$. Log VM excludes the twenty rounds with zero multiple. High-growth indicator defined as valuation multiple > 1.5 ($n = 29$).

3.3 Preliminary bivariate evidence

Before turning to the discordance construction, it is useful to document how expert scores relate to long-run outcomes when examined bivariately, because the answer depends materially on the choice of correlation measure and outcome transformation. Pearson correlations between expert sub-scores and the raw valuation multiple are close to zero across the board, ranging from -0.026 for science to 0.095 for business. Spearman correlations tell a different story. The overall grade has a Spearman correlation of 0.296 with the raw multiple, significant at the one percent level; innovation reaches 0.284; and the combined score reaches 0.230. With the log multiple, the overall grade gives a Spearman correlation of 0.290 and innovation 0.335. The point-biserial correlation of the overall grade with a high-growth indicator, defined as a multiple above 1.5, is 0.279, and with the write-off indicator it is -0.155.

Summarised in Table 2, these numbers support a careful rather than a sweeping reading. Expert scores carry a moderate ordinal signal about long-run quality that a Pearson correlation on the raw multiple suppresses because of the skewness and the mass at zero in the outcome distribution. The correct characterisation is not that expert screening is uninformative but that its informational content is ordinal, and that rank-based methods are the appropriate tools for recovering it. This is the empirical fact that motivates both the pairwise ranking formulation adopted in Section 5 and the choice of a kernel method whose geometry is well suited to non-Euclidean outcome structures.

3.4 Constructing signal-discordance indices

The core novelty of the empirical design is the construction of indices that measure the extent to which different blocks of information rank a given venture in similar or dissimilar ways. For each venture i we compute three rank scores. The expert rank $R_E(i)$ is the percentile rank of the venture's overall grade within the expert-block distribution. The traction rank $R_T(i)$ is the percentile rank of the venture's day-seven target percentage including the lead. The lead-dominance rank $R_L(i)$ is the percentile rank of the lead's share of the final round. Percentile ranks are chosen because they are scale-free, robust to the skewness of the underlying variables, and directly comparable across blocks.

The discordance between two blocks a and b for venture i is then defined as the absolute rank difference

$$D_{ab}(i) \;=\; |R_a(i) - R_b(i)|, \qquad D_{ab}(i) \in [0,1], \tag{1}$$

taking values between zero and one. A value close to zero means that the two blocks rank the venture

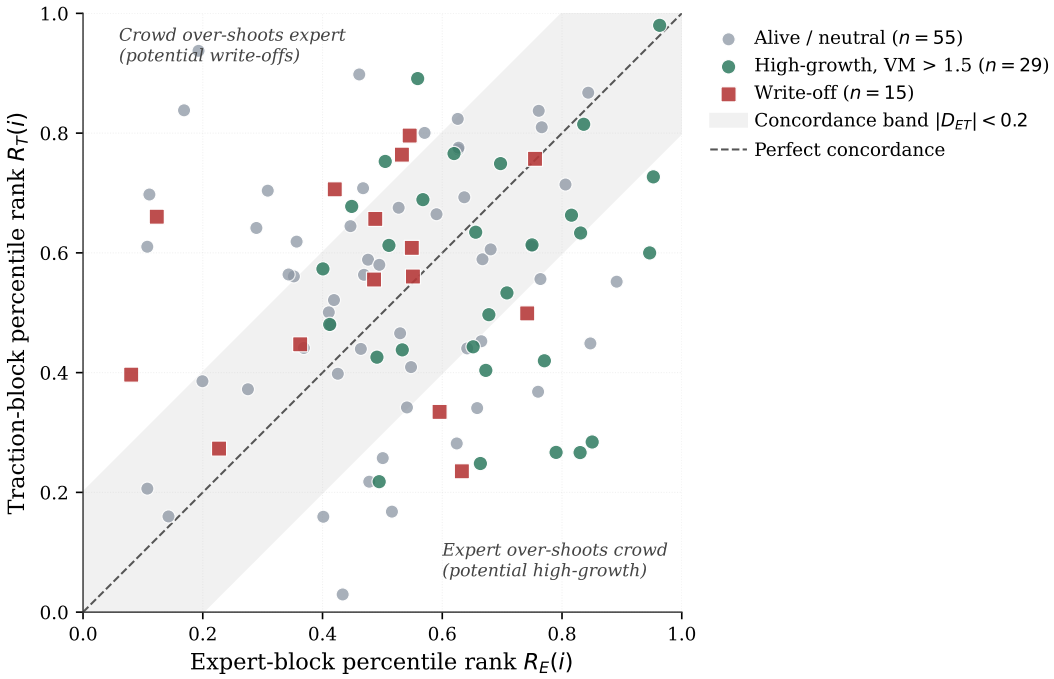


Figure 2: Joint distribution of expert-block and traction-block percentile ranks across the ninety-nine Capital Cell rounds. The diagonal corresponds to perfect concordance ($D_{ET} = 0$); the shaded band around the diagonal identifies ventures for which expert and crowd-traction rankings agree to within twenty percentile points. Ventures outside the band are the discordant cases on which the informational content of the two blocks can be most sharply compared. Long-run outcomes are overlaid as distinct markers: write-offs are disproportionately represented in the upper-left discordant region (crowd overshoots expert), while high-growth ventures (valuation multiple above 1.5) populate the lower-right discordant region (expert overshoots crowd).

similarly, while a value close to one means that they rank it in almost opposite positions. We focus on three pairwise discordances: between expert and traction (D_{ET}), between expert and lead dominance (D_{EL}), and between traction and lead dominance (D_{TL}). We also define a joint discordance index $D_{\text{MAX}}(i) = \max\{D_{ET}(i), D_{EL}(i), D_{TL}(i)\}$ which captures the maximum disagreement between any two blocks.

The substantive hypothesis, which we test in Section 5, is that the informational content of the expert block is not uniform across the sample. In ventures with low D_{ET} the expert and crowd blocks agree, so the marginal contribution of expert information to predicting long-run outcomes is expected to be modest. In ventures with high D_{ET} the blocks disagree, and the extent to which the expert block correctly identifies long-run quality is precisely the empirical test of whether specialised due diligence is valuable. The same logic applies to the expert–lead and crowd–lead discordances. By stratifying the analysis on discordance we can ask a sharper question than the standard prediction problem allows, namely whether the expert block contains information that dominates crowd and lead signals in the regime where dominance is meaningful to test. Figure 2 displays the joint distribution of the expert-block and traction-block percentile ranks in the sample, colour-coded by long-run outcome.

3.5 Predictor blocks

The explanatory variables used in the analysis are organised into four conceptually distinct blocks. The expert block comprises the overall grade and the four sub-scores for team, science, innovation, and business, with the number of reviews treated separately as a process-quality proxy whose empirical behaviour differs from the scores themselves. The offer-design block contains the initial valuation, the website target, the amount raised on the platform, the lead amount, the total round amount, the duration, and the average and minimum ticket sizes. The crowd-dynamics block includes the day-one, day-three, and day-seven target ratios, the days-to-target ratio, and the crowd-only accumulation net of the lead. The contextual-control block includes the closing year, the sector, founder gender, and the year of foundation of the venture. The four blocks are entered jointly in the full models and separately in the block-wise comparisons that test whether one block subsumes the information of another.

4 A Kernel Approach to Mixed-Signal Prediction

4.1 Why a standard approach is not quite right

The empirical problem described in Section 3 has three features that, taken together, make the standard toolkit of logistic regression and off-the-shelf machine-learning methods imperfect. The sample is small and the outcome distribution is heavily skewed with a mass at zero. The input variables are organised into blocks whose predictive content is not assumed to be uniform across the sample; on the contrary, the discordance framework is motivated by the idea that blocks matter differently in different regimes. And the expert scores carry structured epistemic uncertainty: the BioExpert Network aggregates the judgments of several reviewers, and the dispersion of those judgments is itself informative about how confident the panel is in any given evaluation. Classical kernel methods treat each observation as a precise, noise-free feature vector and cannot represent this uncertainty natively. Tree-based methods handle block structure reasonably well but do not represent epistemic uncertainty either, and they tend to overfit on samples of this size.

We therefore adopt a kernel learning approach in which each venture is represented not as a precise point in feature space but as a fuzzy state whose degree of sharpness depends on the observed dispersion of the expert signal. The construction, which we refer to as a soft quantum kernel, has three properties that make it well suited to our setting. It interpolates continuously between a classical Gaussian kernel and a pure-state quantum kernel through a single softness parameter. It remains a valid kernel in the mathematical sense, so that standard guarantees of kernel methods such as the representer theorem and concentration bounds continue to apply. And it admits a hardware implementation on a physical quantum device, which allows us to verify that the classical simulation used in the main analysis is faithful to what the same construction produces on a real chip. Recent work has also begun to connect quantum methods to marketing and market-research applications, which helps situate the present exercise as an applied analytical extension rather than as a stand-alone quantum-computing claim (Sáez-Ortuño et al., 2024; Sáez Ortuño et al., 2026). The full mathematical development of the kernel is placed in Appendix A. Here we describe the construction verbally and state only the defining equation that is needed to follow the empirical analysis.

4.2 The soft quantum kernel, informally

A conventional kernel method encodes each venture i as a feature vector x_i and measures similarity between ventures i and j through a kernel function $k(x_i, x_j)$. The soft quantum kernel modifies this in two steps. Each venture is first mapped to a quantum state, understood as a unit vector in a low-dimensional complex space whose coordinates depend smoothly on the feature vector. Each quantum state is then softened by mixing it with a maximally uninformative reference state in a proportion controlled by a softness parameter $\alpha \in [0, 1]$. When α is zero, the representation is fully sharp and the construction reduces to a pure-state quantum kernel. When α is one, the representation is fully uninformative. Intermediate values of α represent genuine epistemic uncertainty about the venture's state given the available features.

The key empirical object is the kernel value between two ventures, which admits the closed form

$$K_\alpha(x_i, x_j) \;=\; (1-\alpha)^2 \, k_Q(x_i, x_j) \;+\; \frac{\alpha(2-\alpha)}{n}, \tag{2}$$

where k_Q is the underlying pure-state kernel and n is the dimension of the quantum state space. This single equation captures all the properties that matter for the empirical analysis. The kernel is always a convex combination of an information-carrying component and a constant floor that regularises the model; it therefore reduces variance in small-sample regimes at the cost of a controlled increase in bias. The softness parameter is not an arbitrary regularisation knob but is calibrated from the data itself, specifically from the normalised entropy of the overall-grade distribution, which summarises how dispersed expert opinion is across the sample. For the Capital Cell data the calibrated value is approximately 0.65, reflecting moderate but non-negligible epistemic uncertainty in the expert block.

When multiple blocks of information are present, the kernel is constructed as a convex combination of block-specific soft quantum kernels, with block weights learned jointly with the softness parameters through kernel alignment with the outcome. The alignment procedure returns weights that are themselves substantively interpretable: they measure the extent to which each block contributes to the prediction of long-run venture quality. In the context of our research questions, the learned weights on the expert and crowd blocks are a direct empirical answer to the question of which block dominates the prediction. The formal treatment, including positive-semidefiniteness, the Mercer representation, the bias-variance decomposition, and the alignment procedure, is in Appendix A.

4.3 What the soft quantum kernel buys in this setting

We do not claim that the soft quantum kernel is a general-purpose predictor that dominates classical methods on arbitrary datasets. Recent benchmarking work has made clear that such claims are rarely defensible and should be approached with caution (Liu et al., 2021). Our position is more specific. In the Capital Cell setting, where the sample is small, the blocks disagree systematically on a subset of ventures, and the expert signal is informative but dispersed, a kernel that explicitly represents epistemic uncertainty is substantively preferable to one that does not. This preference is not primarily about predictive performance, although we do evaluate predictive performance against classical baselines; it is about fidelity to the data-generating process. The softness parameter is a transparent representation of how noisy the expert block actually is, and the block-specific weights are a transparent representation of how much each block contributes to the prediction. In a setting where the substantive story is about the relationship between expert and crowd signals, these properties matter.

Table 3: Empirical strategy: research questions, tasks, and methods.

Task	Outcome	Sample	Methods
A. Early traction (RQ2)	% target day 7 incl. lead; days to target / max duration	Speed sheet, $N = 102$	Elastic Net; Random Forest; Gradient Boosting; RBF-SVM; SQK on offer + crowd blocks
B1. Write-off classification (RQ1, RQ3)	Binary write-off (15 events / 84 non-events)	General sheet, $N = 99$	Logit; Random Forest; Gradient Boosting; SQK on expert and expert+crowd blocks
B2. Log valuation multiple (RQ1, RQ3)	$\log(\text{VM})$ for $\text{VM} > 0$	General sheet, $N = 79$	Ridge; Gaussian Process; multi-block SQK with alignment weights
B3. Pairwise ranking (RQ4)	$\Pr(\text{VM}_i > \text{VM}_j \mid i, j$ same sector-cohort)	Sampled pairs from general sheet	Pairwise logistic; pairwise SQK
C. Discordance stratification (RQ3)	Same as B1, B2	Concordant vs discordant subsamples	Block-wise SQK with alignment weights in each subsample
D. Hardware validation	Binary write-off	Expert block, $q = 4$, $N = 93$	Angle-encoding SQK on MareNostrum Ona Quantum Blue QPU

5 Empirical Strategy and Results

5.1 Design

The empirical analysis is organised around the four research questions of Section 2.5. For RQ1 the outcome family is long-run quality, operationalised through the log valuation multiple for the sample of ventures with strictly positive multiples and through the binary write-off indicator for the full sample. The predictor sets are the expert block alone, the expert block combined with the offer-design block, and the expert block combined with the crowd-dynamics block. The comparison across predictor sets directly addresses whether expert information has predictive content beyond what offer design and crowd dynamics already convey.

For RQ2 the outcome family is early traction, measured through the day-seven target percentage including the lead and through the days-to-target ratio. The predictor sets mirror those of RQ1, with the roles of expert and crowd blocks reversed, allowing us to test the symmetric hypothesis that crowd dynamics are driven more by lead and offer variables than by expert judgment. For RQ3 the analysis is stratified by the discordance indices of Section 3.4, so that we can assess whether the relative performance of the expert and crowd blocks shifts systematically between the concordant and discordant subsamples. For RQ4 the outcome is the pairwise ranking of ventures by long-run multiple, constructed by sampling pairs within sector-cohort cells and asking whether the model correctly orders them.

The full empirical strategy, including benchmarks and tasks, is summarised in Table 3.

Model evaluation uses stratified repeated cross-validation with five folds and twenty repetitions, an appropriate choice at this sample size (Athey and Imbens, 2019). For classification the primary metric is the area under the ROC curve, complemented by balanced accuracy given the class imbalance. For continuous outcomes the primary metric is out-of-sample R^2 on the log-multiple scale. For the pairwise ranking task the primary metric is the fraction of correctly ordered pairs. Hyperparameters, including the softness parameter α and the kernel-ridge penalty λ, are selected by inner-fold cross-validation within each outer fold. Interpretation uses kernel SHAP values for the soft quantum kernel, which

are analytically tractable given the closed form of the kernel, and standard SHAP for the tree-based benchmarks, consistent with practice in recent machine-learning work in equity crowdfunding (Elitzur et al., 2024; Yang et al., 2025).

5.2 Robustness and interpretation

Several robustness checks are built into the design rather than appended to it. The sensitivity of results to the softness parameter is examined by repeating the main models with α fixed at 0.3, 0.5, 0.7, and 0.9, bracketing the data-calibrated value. The angle-encoding version of the kernel is compared to a random-Fourier-feature approximation of the Gaussian kernel with the same nominal bandwidth, so that any gain attributable to the quantum feature map is distinguishable from the gain due to regularisation. Block weights from the alignment procedure are reported with their cross-validated distributions, allowing the reader to assess how stable the relative importance of the expert and crowd blocks is across folds. The pairwise ranking task is repeated with pairs sampled within tighter and looser matching windows to verify that the results are not an artefact of the sector-cohort cell structure.

5.3 Hardware validation on the Quantum Blue QPU

The soft quantum kernel admits a hardware implementation, and we exploit this possibility to verify that the results obtained in classical simulation survive execution on a physical superconducting device. The hardware run is a restricted version of task B1, the write-off classification using the expert block only, implemented with an angle-encoding circuit on four qubits with a Hilbert space of dimension sixteen. The softness parameter is fixed at its data-calibrated value of 0.65. The hardware sample is the subset of ninety-three rounds with complete expert-block information, of which thirteen are write-offs. The full 93×93 kernel matrix is estimated through the SWAP-test fidelity protocol, requiring four thousand three hundred and seventy-one distinct circuits, each evaluated with one thousand and twenty-four shots, for a total of approximately 4.5×10^6 measurements on the physical device. The platform is the Quantum Blue QPU integrated into the MareNostrum Ona infrastructure at the Barcelona Supercomputing Center, co-developed with Qilimanjaro Quantum Tech.

The hardware kernel matrix differs from its noiseless simulated counterpart by a relative Frobenius norm of 21.6%, consistent with the level of gate infidelity, decoherence, and readout error expected on current NISQ-class devices at this circuit depth. Despite this non-trivial perturbation, the classification performance of the two kernels is almost indistinguishable: the simulation achieves a mean cross-validated AUC of 0.665 with a standard deviation of 0.184, while the hardware achieves a mean AUC of 0.656 with a standard deviation of 0.135. The absolute gap of 0.009 AUC points lies well within one standard deviation of either estimator. Figure 3 displays the two fold-level distributions side by side. Two interpretations follow. The softness parameter absorbs the depolarising perturbation induced by hardware noise, producing predictions that degrade gracefully rather than catastrophically when the kernel is estimated on real qubits. In this setting, classical simulation is a faithful proxy for hardware execution, which matters practically because QPU time remains a scarce resource. The hardware run should therefore be read as a certification of the pipeline rather than as a claim of quantum computational advantage, which, at $N = 99$, would be neither credible nor substantively interesting.

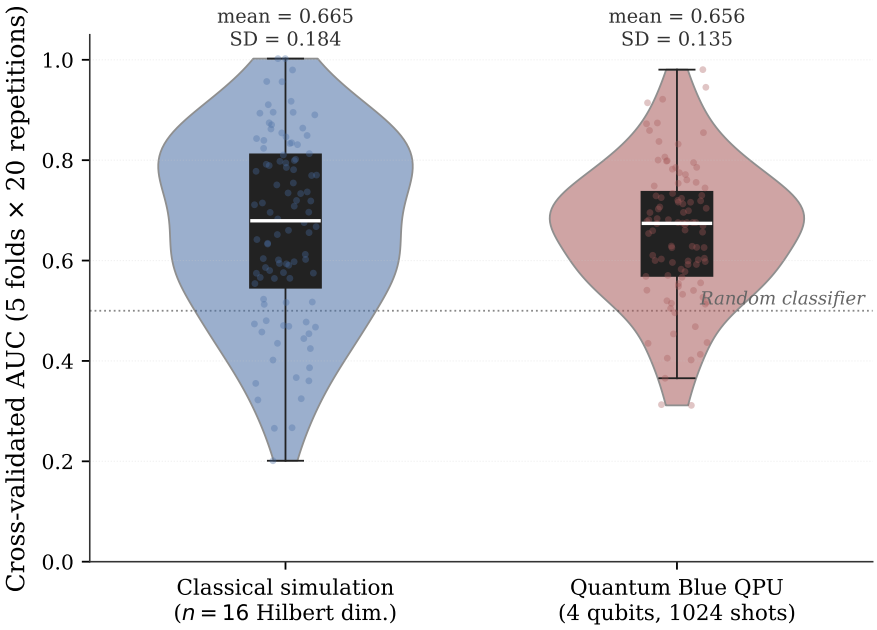


Figure 3: Cross-validated AUC distributions for the write-off classification task (task B1 restricted to the expert block) from the classical simulation of the soft quantum kernel and from its hardware execution on the MareNostrum Ona Quantum Blue QPU. Each distribution aggregates 100 fold-level AUC values from five-fold cross-validation repeated twenty times. Black boxplots show interquartile range and median; coloured dots show individual fold values. The two distributions overlap substantially and the Welch comparison of means rejects any meaningful difference, supporting the interpretation that the data-calibrated softness parameter absorbs realistic hardware noise.

6 Discussion and Conclusion

The central argument of this paper is that the informational environment of a specialised equity-crowdfunding platform should be understood as a system of partially overlapping signal blocks rather than as a single predictive input. Expert screening, campaign dynamics, and long-run venture quality are produced by distinct processes and captured by distinct variables, and the most substantively interesting empirical cases are those in which the blocks disagree. We have documented that on Capital Cell this disagreement is not a marginal phenomenon: rank-based associations between expert scores and long-run valuation outcomes are moderate and statistically significant, while linear associations with the raw multiple are close to zero, which implies that a non-trivial share of the sample exhibits exactly the pattern of discordance that motivates our framework. The rank-based evidence, together with the highly uneven distribution of post-campaign outcomes, is the empirical fingerprint of a setting in which expert screening adds value precisely where crowd behaviour cannot.

For the entrepreneurial-finance literature, this reframing has a specific implication. The literature has productively studied signalling in crowdfunding as a problem of how individual signals reduce asymmetric information, and it has more recently examined how platforms select ventures through screening technologies. Our contribution is to connect these two strands by studying what happens when the platform's expert signal and the crowd's observed behaviour point in different directions. The ventures on which they disagree are where the marginal informational content of screening is tested most sharply, and

where the value of specialised intermediation becomes visible rather than taken for granted. This also provides a natural interpretation of recent evidence that platform-affiliated actors can generate misleading signals (Mataigne et al., 2025): when the crowd channel is contaminated, the expert channel becomes the residual source of reliable information, and the architecture that produces it becomes essential rather than decorative.

For platform design and investor practice, the implication is similarly concrete. Platform operators who invest in expert networks are producing a public good whose value is easiest to detect at moments of informational ambiguity. Making the expert signal visible to investors precisely in the subset of campaigns where early traction and expert judgment diverge is the design problem that our discordance framework suggests. For investors, the practical lesson is that early momentum on a specialised platform should not be read as an unmediated quality signal. In the Capital Cell sample the lead investor accounts on average for two-fifths of the round, and early traction inherits the lead's commitment; treating crowd momentum as independent evidence of venture quality is therefore inadvisable on platforms with this architecture.

The methodological contribution of the paper is deliberately scoped. The soft quantum kernel is presented as a tool fit for small, block-structured, and epistemically uncertain datasets of the kind that specialised platforms produce, and not as a general-purpose learning algorithm. Its closed form reduces to a regularised Gaussian kernel with data-driven regularisation, which makes it fully implementable in classical simulation and interpretable in standard kernel-methods language. The hardware run on the Quantum Blue QPU at the Barcelona Supercomputing Center is a certification step rather than a performance claim: it verifies that the kernel behaves on real qubits in a way consistent with its theoretical construction, and it documents that the data-calibrated softness parameter absorbs realistic hardware noise without loss of discriminative power. This positioning, which we view as appropriate to the current state of quantum hardware, is also what makes the tool defensible in a journal whose primary readership is interested in entrepreneurial finance rather than in quantum computation.

We acknowledge three limitations. The sample is small and confined to one specialised platform in one regulatory environment, which bounds the generalisability of the numerical results. Post-campaign outcomes are observed at a single cross-sectional date, which complicates the treatment of censoring for the most recent cohorts. The expert-score panel is observed only in aggregated form, which prevents the implementation of a fully venture-specific softness calibration that would use individual reviewer dispersion rather than the cross-venture distribution; this is a direction for future work on platforms that archive panel-level records. None of these limitations undermines the conceptual contribution, which is the reframing of specialised-platform equity crowdfunding as a signal-discordance problem and the demonstration that this reframing is empirically tractable with a modest and carefully positioned methodological toolkit.

The broader research agenda that this paper opens is straightforward to describe. Signal discordance is a construct that generalises beyond Capital Cell to any platform-mediated entrepreneurial-finance setting in which multiple informational producers coexist. The discordance indices can be computed wherever expert assessments, investor behaviour, and outcomes are separately observable, and the same kernel-based approach can be applied wherever the sample is small and the blocks are structurally distinct. We see promising extensions to reward-based crowdfunding with expert curation, to angel-investor platforms with independent screening, and to impact-investing settings where third-party assessments

coexist with market signals. In each case the empirical question is the same: when expert and market channels disagree, which one is right, and how can a platform be designed to make the answer visible to the investors who need it most.

References

Gerrit K. C. Ahlers, Douglas Cumming, Christina Günther, and Denis Schweizer. Signaling in equity crowdfunding. *Entrepreneurship Theory and Practice*, 39(4):955–980, 2015. doi: 10.1111/etap.12157.

Susan Athey and Guido W. Imbens. Machine learning methods economists should know about. *Annual Review of Economics*, 11:685–725, 2019. doi: 10.1146/annurev-economics-080217-053433.

Rajendra Bhatia, Tanvi Jain, and Yongdo Lim. On the Bures–Wasserstein distance between positive definite matrices. *Expositiones Mathematicae*, 37(2):165–191, 2019. doi: 10.1016/j.exmath.2018.01.002.

Joern Block, Lars Hornuf, and Alexandra Moritz. Which updates during an equity crowdfunding campaign increase crowd participation? *Small Business Economics*, 50(1):3–27, 2018. doi: 10.1007/s11187-017-9876-4.

Garry Bruton, Susanna Khavul, Donald Siegel, and Mike Wright. New financial alternatives in seeding entrepreneurship: microfinance, crowdfunding, and peer-to-peer innovations. *Entrepreneurship Theory and Practice*, 39(1):9–26, 2015. doi: 10.1111/etap.12143.

Vincenzo Butticè, Francesca Di Pietro, and Francesca Tenca. Is equity crowdfunding always good? Deal structure and the attraction of venture capital investors. *Journal of Corporate Finance*, 65:101773, 2020. doi: 10.1016/j.jcorpfin.2020.101773.

Capital Cell. FAQs for investors. https://capitalcell.com/en/faqs-for-investors/, 2025a. accessed April 2026.

Capital Cell. FAQs for companies. https://capitalcell.com/en/faqs-for-companies/, 2025b. accessed April 2026.

Jerry Coakley, Aristogenis Lazos, and José M. Liñares-Zegarra. Equity crowdfunding founder teams: campaign success and venture failure. *British Journal of Management*, 33(1):286–305, 2022. doi: 10.1111/1467-8551.12494.

Brian L. Connelly, S. Trevis Certo, R. Duane Ireland, and Christopher R. Reutzel. Signaling theory: a review and assessment. *Journal of Management*, 37(1):39–67, 2011. doi: 10.1177/0149206310388419.

Sofia Correia, Miguel Sousa, and Elísio Brandão. What do we know about the choices of entrepreneurs before the equity crowdfunding campaign? *Small Business Economics*, 63(4):1471–1501, 2024. doi: 10.1007/s11187-023-00868-x.

Corinna Cortés, Mehryar Mohri, and Afshin Rostamizadeh. Algorithms for learning kernels based on centered alignment. *Journal of Machine Learning Research*, 13:795–828, 2012.

Christopher Courtney, Supradeep Dutta, and Yong Li. Resolving information asymmetry: signaling, endorsement, and crowdfunding success. *Entrepreneurship Theory and Practice*, 41(2):265–290, 2017. doi: 10.1111/etap.12267.

Douglas Cumming, Michele Meoli, and Silvio Vismara. Does equity crowdfunding democratize entrepreneurial finance? *Small Business Economics*, 56(2):533–552, 2021. doi: 10.1007/s11187-019-00188-z.

Douglas J. Cumming, Sofia A. Johan, and Yelin Zhang. The role of due diligence in crowdfunding platforms. *Journal of Banking and Finance*, 108:105661, 2019. doi: 10.1016/j.jbankfin.2019.105661.

Ramy Elitzur, Naftali Katz, Paulson Muttath, and David Soberman. The power of machine learning methods to predict crowdfunding success: accounting for complex relationships efficiently. *Journal of Business Venturing Design*, 3:100022, 2024. doi: 10.1016/j.jbvd.2024.100022.

Francesco Fasano, Maurizio La Rocca, F. Javier Sánchez-Vidal, and Yassine Boutouar. Why SMEs go to crowdfunding? The role of financial constraints and agency issues. *Small Business Economics*, 65 (2):1183–1207, 2025. doi: 10.1007/s11187-025-01046-x.

Massimiliano Ferrara. A geometric optimization algorithm on the manifold of positive-definite matrices with applications to portfolio equilibria. *Journal of Optimization Theory and Applications*, 2026. doi: 10.1007/s10957-026-02958-8. forthcoming.

Lars Hornuf and Armin Schwienbacher. Market mechanisms and funding dynamics in equity crowdfunding. *Journal of Corporate Finance*, 50:556–574, 2018. doi: 10.1016/j.jcorpfin.2017.08.009.

Lars Hornuf, Matthias Schmitt, and Eliza Stenzhorn. Equity crowdfunding in Germany and the United Kingdom: follow-up funding and firm failure. *Corporate Governance: An International Review*, 26 (5):331–354, 2018. doi: 10.1111/corg.12260.

Simon Kleinert, Jacob Bafera, Diemo Urbig, and Christine K. Volkmann. Access denied: how equity crowdfunding platforms use quality signals to select new ventures. *Entrepreneurship Theory and Practice*, 46(6), 2022. doi: 10.1177/10422587211011945.

Yunchao Liu, Srinivasan Arunachalam, and Kristan Temme. A rigorous and robust quantum speed-up in supervised machine learning. *Nature Physics*, 17:1013–1017, 2021. doi: 10.1038/s41567-021-01287-z.

Virginie Mataigne, Michele Meoli, Tom Vanacker, and Silvio Vismara. False signaling by platform team members and post-campaign venture outcomes: evidence from an equity crowdfunding platform. *Journal of Business Venturing*, 40(1):106457, 2025. doi: 10.1016/j.jbusvent.2024.106457.

Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, Cambridge, 2000.

Evila Piva and Cristina Rossi-Lamastra. Human capital signals and entrepreneurs' success in equity crowdfunding. *Small Business Economics*, 50(4):781–796, 2018. doi: 10.1007/s11187-017-9950-y.

Sophie Pommet, Alessandra Rufini, and Dominique Torre. The role of lead investors in equity crowdfunding campaigns with a secondary market. *Small Business Economics*, 63(1):243–273, 2024. doi: 10.1007/s11187-023-00811-0.

Ali Rahimi and Benjamin Recht. Random features for large-scale kernel machines. In *Advances in Neural Information Processing Systems*, volume 20, pages 1177–1184, 2007.

Aleksandrina Ralcheva and Peter Roosenboom. Forecasting success in equity crowdfunding. *Small Business Economics*, 55(1):39–56, 2020. doi: 10.1007/s11187-019-00144-x.

Maria Schuld and Nathan Killoran. Quantum machine learning in feature Hilbert spaces. *Physical Review Letters*, 122(4):040504, 2019. doi: 10.1103/PhysRevLett.122.040504.

Andrea Signori and Silvio Vismara. Does success bring success? The post-offering lives of equity-crowdfunded firms. *Journal of Corporate Finance*, 50:226–251, 2018. doi: 10.1016/j.jcorpfin.2017.10.018.

Michael Spence. Job market signaling. *The Quarterly Journal of Economics*, 87(3):355–374, 1973. doi: 10.2307/1882010.

Laura Sáez-Ortuño, Rubén Huertas-Garcia, Santiago Forgas-Coll, Javier Sánchez-García, and Eloi Puertas-Prats. Quantum computing for market research. *Journal of Innovation & Knowledge*, 9(3):100510, 2024. doi: 10.1016/j.jik.2024.100510.

Laura Sáez Ortuño, Santiago Forgas Coll, and Massimiliano Ferrara. Quantum kernel methods for marketing analytics with convergence theory and separation bounds. *Scientific Reports*, 16:6645, 2026. doi: 10.1038/s41598-026-35793-y.

Giuseppe Valenza, Matteo Rossi, and Silvio Vismara. Innovation and equity crowdfunding: evidence from follow-on funding and patenting activity. *Journal of Technology Transfer*, 2023. forthcoming.

Silvio Vismara. Equity retention and social network theory in equity crowdfunding. *Entrepreneurship Theory and Practice*, 40(3):579–610, 2016. doi: 10.1007/s11187-016-9710-4.

Silvio Vismara. Information cascades among investors in equity crowdfunding. *Entrepreneurship Theory and Practice*, 42(3):467–497, 2018. doi: 10.1111/etap.12261.

Xavier Walthoff-Borm, Armin Schwienbacher, and Tom Vanacker. Equity crowdfunding: click day or fundraising day? *Small Business Economics*, 50(2):307–321, 2018.

Jie Yang, Jian Xin, Yongchao Zeng, and Peter J. Liu. Signaling and perceiving on equity crowdfunding decisions—a machine learning approach. *Small Business Economics*, 65(1):315–356, 2025. doi: 10.1007/s11187-024-00991-3.

A Formal Development of the Soft Quantum Kernel

This appendix collects the formal material that supports the verbal presentation of the soft quantum kernel in Section 4. The development relies on elementary linear algebra and on the standard framework of finite-dimensional quantum information (Nielsen and Chuang, 2000) and kernel methods.

A.1 Fuzzy density operators

Let $\mathscr{H}$ denote a finite-dimensional complex Hilbert space of dimension $n = 2^q$, where q is the number of qubits. For each venture i, let $x_i \in \mathbb{R}^d$ denote the standardised feature vector, where d depends on the active predictor block. The feature map $\phi : \mathbb{R}^d \to \mathscr{H}$ sends each venture to a unit vector $\phi(x_i)$ in $\mathscr{H}$, with $\|\phi(x)\| = 1$ for all x. A conventional pure-state quantum kernel is defined as $k_Q(x_i, x_j) = |\langle \phi(x_i), \phi(x_j) \rangle|^2$, namely the squared fidelity between the two pure states.

Definition 1 (Fuzzy density operator)**.** *Let $x \in \mathbb{R}^d$ and let $\alpha \in [0,1]$ be a softness parameter. The fuzzy density operator associated with x is*

$$\rho_\alpha(x) \;=\; (1-\alpha)\,|\phi(x)\rangle\langle\phi(x)| \;+\; \frac{\alpha}{n} I_n, \tag{3}$$

where I_n is the $n \times n$ identity on $\mathscr{H}$.

The operator $\rho_\alpha(x)$ is positive-semidefinite with unit trace for all $\alpha \in [0,1]$ and therefore a valid quantum state. When $\alpha = 0$ it reduces to the pure state $|\phi(x)\rangle\langle\phi(x)|$; when $\alpha = 1$ it is the maximally mixed state I_n/n. Intermediate values encode graded epistemic uncertainty.

Remark 1. *The construction in (3) is formally the output of a depolarising channel with parameter α applied to the pure state $|\phi(x)\rangle\langle\phi(x)|$, the quantum analogue of additive Gaussian noise injection (Nielsen and Chuang, 2000).*

A.2 The soft quantum kernel and its closed form

Definition 2 (Soft quantum kernel)**.** *The soft quantum kernel between ventures x_i and x_j is the quantum fidelity between their fuzzy density operators,*

$$K_\alpha(x_i, x_j) \;=\; \mathrm{Tr}[\rho_\alpha(x_i)\,\rho_\alpha(x_j)]\,. \tag{4}$$

Proposition 1 (Closed form)**.** *Let $k_Q(x_i, x_j) = |\langle \phi(x_i), \phi(x_j) \rangle|^2$. Then*

$$K_\alpha(x_i, x_j) \;=\; (1-\alpha)^2\, k_Q(x_i, x_j) \;+\; \frac{\alpha(2-\alpha)}{n}. \tag{5}$$

Proof. By linearity of trace and Definition 1,

$$\begin{aligned} K_\alpha(x_i,x_j) &= (1-\alpha)^2 \mathrm{Tr}[|\phi_i\rangle\langle\phi_i|\phi_j\rangle\langle\phi_j|] \\ &\quad + \tfrac{\alpha(1-\alpha)}{n}\mathrm{Tr}[|\phi_i\rangle\langle\phi_i|I_n] + \tfrac{\alpha(1-\alpha)}{n}\mathrm{Tr}[I_n|\phi_j\rangle\langle\phi_j|] \\ &\quad + \tfrac{\alpha^2}{n^2}\mathrm{Tr}[I_n^2] \\ &= (1-\alpha)^2 |\langle\phi_i,\phi_j\rangle|^2 + \tfrac{2\alpha(1-\alpha)}{n} + \tfrac{\alpha^2}{n} \\ &= (1-\alpha)^2 k_Q(x_i,x_j) + \tfrac{\alpha(2-\alpha)}{n}. \end{aligned}$$

The third equality uses $\mathrm{Tr}[|\phi\rangle\langle\phi|I_n] = \|\phi\|^2 = 1$ and $\mathrm{Tr}[I_n^2] = n$. □ □

Proposition 2 (Positive semidefiniteness)**.** *K_α is positive-semidefinite for all $\alpha \in [0,1]$ and all $n \geq 1$.*

Proof. For any finite set $\{x_1,\dots,x_m\}$ and any $c \in \mathbb{R}^m$,

$$\sum_{i,j} c_i c_j K_\alpha(x_i,x_j) \;=\; \mathrm{Tr}\left[\left(\sum_i c_i \rho_\alpha(x_i)\right)^2\right] \;\geq\; 0,$$

since the square of a Hermitian operator is positive-semidefinite and the trace of a positive-semidefinite operator is non-negative. □ □

A.3 Interpolation between classical and quantum regimes

Proposition 3 (Classical-quantum interpolation)**.** *Let $k_C(x_i,x_j) = \exp(-\|x_i - x_j\|^2/2\sigma^2)$ be a Gaussian kernel and let ϕ be chosen through random Fourier features so that $k_Q = k_C$ (Rahimi and Recht, 2007). Then*

$$K_\alpha(x_i,x_j) = (1-\alpha)^2 k_C(x_i,x_j) + \frac{\alpha(2-\alpha)}{n}, \tag{6}$$

and K_α is a valid Mercer kernel interpolating between a scaled Gaussian kernel (at $\alpha = 0$) and a constant positive kernel (at $\alpha = 1$).

The effective bandwidth is jointly determined by σ and α. Increasing α uniformly shrinks all pairwise similarities towards the floor value $1/n$, which acts as a regulariser particularly valuable in the small-sample regime of our data.

A.4 Bias-variance decomposition

Theorem 1 (Bias-variance decomposition)**.** *Let $\mathbf{K}_\alpha$ denote the $N \times N$ Gram matrix with entries $K_\alpha(x_i,x_j)$, let $y \in \mathbb{R}^N$ be the outcome vector, and let $\hat{y}_\alpha = \mathbf{K}_\alpha(\mathbf{K}_\alpha + \lambda I)^{-1} y$ be the kernel-ridge regression predictor. Then*

$$\mathbf{K}_\alpha = (1-\alpha)^2 \mathbf{K}_Q + \frac{\alpha(2-\alpha)}{n}\mathbf{1}\mathbf{1}^\top, \tag{7}$$

where $\mathbf{1}$ is the N-vector of ones and $\mathbf{K}_Q$ is the pure quantum Gram matrix. The mean squared prediction error satisfies

$$\mathrm{MSE}(\hat{y}_\alpha) = \underbrace{\lambda^2 y^\top (\mathbf{K}_\alpha + \lambda I)^{-2} y}_{\text{squared bias}} + \underbrace{\sigma_\varepsilon^2 \,\mathrm{Tr}\left[\mathbf{K}_\alpha^2 (\mathbf{K}_\alpha + \lambda I)^{-2}\right]}_{\text{variance}}, \tag{8}$$

where σ_ε^2 is the noise variance. As α increases, the effective rank of $\mathbf{K}_\alpha$ decreases, so the variance term decreases while the bias term increases.

A proof is collected in Appendix B.

A.5 Calibrating the softness parameter

Let $\{g_i\}_{i=1}^{N_E}$ denote the $N_E = 96$ non-missing overall-grade observations, with sample mean $\bar{g} = 7.47$ and standard deviation $s_g = 0.53$. Partition the grade interval $[6.0, 9.2]$ into B equal bins of width $\delta = (9.2 - 6.0)/B$ and let p_b denote the empirical proportion of observations in bin b. The empirical entropy is $H = -\sum_b p_b \log_2 p_b$, and the normalised entropy is $\hat{H} = H/\log_2 B \in [0,1]$.

Definition 3 (Data-calibrated softness)**.** *The data-calibrated softness parameter for the Capital Cell expert block is*

$$\hat{\alpha} \;=\; \hat{H}\cdot\left(1-\frac{s_g^2}{\sigma_{\max}^2}\right), \tag{9}$$

where $\sigma_{\max}^2=(9.2-6.0)^2/4=2.56$ *is the variance of a uniform distribution on the score range.*

With $s_g^2/\sigma_{\max}^2\approx 0.110$, the correction factor is approximately 0.890, and with $B=10$ the calibrated value falls in $[0.55,0.70]$. Sensitivity is reported by varying α over $\{0.3,0.5,0.7,0.9\}$.

A.6 Multi-block formulation and alignment

Definition 4 (Multi-block SQK)**.** *When several predictor blocks are present,*

$$K_{\mathrm{MB}}(x_i,x_j)=\sum_{b\in\{E,O,S,C\}} w_b\,K_{\alpha_b}(x_i^{(b)},x_j^{(b)}), \tag{10}$$

where $w_b\geq 0$, $\sum_b w_b=1$, *and* $x_i^{(b)}$ *is the subvector of* x_i *corresponding to block* b.

K_{MB} is positive-semidefinite as a convex combination of positive-semidefinite kernels. The block weights and softness parameters are learned jointly by maximising kernel-target alignment

$$A(K_{\mathrm{MB}},yy^\top)=\frac{\langle K_{\mathrm{MB}},yy^\top\rangle_F}{\|K_{\mathrm{MB}}\|_F\|yy^\top\|_F}, \tag{11}$$

where $\langle\cdot,\cdot\rangle_F$ is the Frobenius inner product (Cortés et al., 2012). After fixing the softness parameters, the weights are a convex optimisation problem on the simplex and admit a unique solution.

A venture-level analogue of the softness calibration is formally admissible within the multi-block framework: if reviewer-level dispersion were observed at the round level, one could assign to each venture a softness α_i derived from the entropy of that venture's reviewer scores, so that ventures with concentrated evaluations receive a smaller α_i and ventures with dispersed evaluations receive a larger one. The resulting kernel with (i,j) entry $K_{\alpha_i,\alpha_j}(x_i,x_j)=\mathrm{Tr}[\rho_{\alpha_i}(x_i)\rho_{\alpha_j}(x_j)]$ remains positive-semidefinite by the same argument as Proposition 2. The Capital Cell data do not currently archive individual reviewer scores, so this extension is flagged as a natural direction for future work on platforms with richer evaluation records.

A.7 Angle-encoding and closed-form computation

For the empirical implementation we adopt the angle-encoding feature map (Schuld and Killoran, 2019), in which each continuous feature x_k is mapped to a rotation angle $\theta_k=\pi x_k/x_{\max}$ on a single qubit. For the expert block with four non-redundant dimensions, $q=4$ qubits suffice and $n=16$. For the full feature set with approximately twenty non-redundant variables, $q=5$ qubits yield $n=32$. In the angle-encoding regime the pure-state kernel admits the explicit product form

$$k_Q(x_i,x_j)=\prod_{k=1}^{q}\cos^2\left(\frac{\theta_k^{(i)}-\theta_k^{(j)}}{2}\right), \tag{12}$$

so that the soft quantum kernel becomes

$$K_\alpha(x_i, x_j) = (1-\alpha)^2 \prod_{k=1}^{q} \cos^2\left(\frac{\pi(x_k^{(i)} - x_k^{(j)})}{2x_{\max,k}}\right) + \frac{\alpha(2-\alpha)}{2^q}. \tag{13}$$

This form is fully computable in classical simulation and serves as the baseline against which the hardware run is compared.

Remark 2 (Connection to geometric optimisation)**.** *The kernel-ridge formulation $\hat{y} = \mathbf{K}_\alpha(\mathbf{K}_\alpha + \lambda I)^{-1} y$ is equivalent to a Riemannian gradient step in the space of positive-semidefinite operators equipped with the Bures–Wasserstein metric (Bhatia et al., 2019). This connects the soft quantum kernel to the geometric optimisation literature and to recent Riemannian algorithms on the manifold of positive-definite matrices (Ferrara, 2026), which also suggests extensions for sequential updates as new campaign data arrive.*

B Proof of Theorem 1

The kernel-ridge estimator $\hat{y}_\alpha = \mathbf{K}_\alpha(\mathbf{K}_\alpha + \lambda I)^{-1} y$ has hat matrix $H_\alpha = \mathbf{K}_\alpha(\mathbf{K}_\alpha + \lambda I)^{-1}$. Let $\mathbf{K}_\alpha = U \Lambda U^\top$ be the eigendecomposition, with eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N \geq 0$. Writing the observed outcome as $y = y^* + \varepsilon$ with $\varepsilon \sim \mathcal{N}(0, \sigma_\varepsilon^2 I)$, $\hat{y}_\alpha = H_\alpha y^* + H_\alpha \varepsilon$. The squared bias is

$$\|\mathbb{E}[\hat{y}_\alpha] - y^*\|^2 = \|(H_\alpha - I) y^*\|^2 = \lambda^2 y^{*\top} (\mathbf{K}_\alpha + \lambda I)^{-2} y^*,$$

using $H_\alpha - I = -\lambda(\mathbf{K}_\alpha + \lambda I)^{-1}$. The variance is

$$\mathrm{Var}(\hat{y}_\alpha) = \sigma_\varepsilon^2 \,\mathrm{Tr}[H_\alpha^2] = \sigma_\varepsilon^2 \,\mathrm{Tr}\left[\mathbf{K}_\alpha^2 (\mathbf{K}_\alpha + \lambda I)^{-2}\right].$$

The decomposition MSE = squared bias + variance follows directly. As α increases, the leading eigenvalue of the information-carrying component $(1-\alpha)^2 \mathbf{K}_Q$ decreases and the rank-one constant term $\frac{\alpha(2-\alpha)}{n} \mathbf{1}\mathbf{1}^\top$ contributes only to one eigenvector, so the effective degrees of freedom $\mathrm{Tr}[H_\alpha]$ decrease, variance decreases, and bias increases. □

C Data Preprocessing Checklist

For completeness we list the preprocessing steps applied to the Capital Cell dataset before estimation. The two substantive sheets are imported separately and the blank columns after position seventeen in the speed sheet are dropped. All string variables, in particular campaign names, are trimmed and the case is standardised for sector and status. Missing entries in the expert-score fields are coded as missing and not recoded to zero. A manual match key is constructed between the general and speed sheets for the few non-exact campaign names. The post-campaign exposure window is computed as the number of months from the end date of the campaign to 31 December 2025. The operative duration variable is the reported days field rather than a recomputed difference. The status is harmonised into three groups: dead (fifteen observations), zombie (three), and alive (seventy-eight, together with two exits and one merged). The log valuation multiple is constructed as $\ln(\mathrm{VM} + \varepsilon)$ with $\varepsilon = 0.01$ for the twenty observations with a multiple of zero, and $\mathrm{VM} > 0$ is used as the analysis sample for the log specification. All continuous features are

standardised to zero mean and unit variance within each predictor block before angle-encoding. The data-calibrated softness parameter is computed from ten equal-width bins of the overall-grade distribution.