

The Semantic Risk Premium

Rouzbeh Rezaei Sanjabi^{1,*} 

¹University of Indianapolis, Indianapolis IN 46227, USA

*Corresponding author: rezaeisanjabir@uindy.edu

Abstract

Prediction markets aggregate information through prices, yet their efficacy depends critically on the clarity of contract language. We introduce the *Semantic Risk Score* (SRS), a novel measure of linguistic ambiguity in prediction market contracts derived from large language model predictions of dispute probability. Using dispute data from the UMA Protocol ($N = 804$) to validate our measure, we demonstrate that SRS correlates significantly with actual oracle dispute variance ($\rho = 0.157$, $p < 0.001$). Applying this validated instrument to Polymarket contracts ($N = 246$), we find that semantic risk manifests through liquidity withdrawal: a one-standard-deviation increase in SRS is associated with a substantial collapse in trading volume ($t = -4.41$, $p < 0.001$). Paradoxically, high-ambiguity markets converge faster to extreme prices, which we attribute to adverse selection: sophisticated traders exit ambiguous markets, leaving homogeneous participants who rapidly reach false consensus. Our findings extend the “Market for Lemons” framework to prediction markets and have implications for contract design, portfolio optimization, and decentralized oracle governance.

Keywords: prediction markets, semantic risk, oracle problem, natural language processing, market microstructure, adverse selection, DeFi

1 Introduction

Prediction markets have emerged as powerful mechanisms for aggregating dispersed information into prices (Hanson, 2003). Analogous to the variance risk premium in equity markets (Bollerslev et al., 2009), we posit the existence of a *semantic risk premium*—a compensation demanded for bearing the risk of linguistic ambiguity. From forecasting elections to anticipating scientific breakthroughs, these markets leverage the wisdom of crowds to produce probability estimates that often outperform expert forecasts. Yet the theoretical elegance of prediction markets rests on an often-overlooked assumption: that contracts are unambiguously interpretable.

In practice, prediction market contracts are specified in natural language, which introduces inherent ambiguity. Consider the question “Will Trump launch a coin before the election?” Reasonable traders might disagree on what constitutes “launching” (announcing? minting? trading?) or what counts as a “coin” (official currency? cryptocurrency? commemorative token?). This linguistic uncertainty creates what we term *semantic risk*: the probability that a contract’s resolution will be disputed due to ambiguous wording rather than uncertain outcomes.

The consequences of semantic risk are particularly acute in decentralized prediction markets, where resolution depends on oracle mechanisms rather than centralized authorities. Platforms like Polymarket rely on external oracles such as UMA Protocol to adjudicate outcomes, and disputes can freeze capital, delay payouts, and undermine market integrity. Despite the practical importance of this phenomenon, no systematic method exists for quantifying semantic risk ex ante. This paper makes three contributions. First, we develop a novel metric, the Semantic Risk Score (SRS), that uses large language models to predict the dispute probability of prediction

market contracts based solely on their textual description. Drawing on recent advances in uncertainty quantification for natural language generation (Kuhn et al., 2023), we operationalize semantic ambiguity as the predicted Bernoulli variance of oracle votes. Second, we validate this measure using ground-truth dispute data from the UMA Protocol, demonstrating that LLM-derived SRS correlates significantly with actual voting variance. Third, we apply this validated instrument to Polymarket to test whether semantic risk affects market behavior.

Our findings challenge conventional intuitions about risk in prediction markets. Semantic risk operates through a liquidity channel: traders systematically avoid ambiguous contracts, resulting in thinner order books and lower trading volume. This pattern is consistent with Akerlof (1970) adverse selection framework: when contract quality (clarity) is difficult to assess, sophisticated traders exit, and markets fail.

A second notable finding emerges from our analysis: high-ambiguity markets converge *faster* to extreme prices (near \$0 or \$1), not slower. We propose a selection-based explanation: when sophisticated traders exit ambiguous markets, the remaining participants are those who either fail to perceive the ambiguity or hold strong prior convictions. This homogeneous population rapidly reaches consensus, but it is a false consensus born of adverse selection rather than genuine information aggregation.

These findings have practical implications for prediction market design and participation. For platform operators, our results suggest that linguistic clarity should be a first-order concern in contract creation, perhaps more important than topic selection or market-making incentives. For traders and market makers, we develop a portfolio optimization framework that incorporates semantic risk as a liquidity factor, demonstrating that filtering high-SRS contracts substantially improves capital efficiency.

2 Related Literature

2.1 Prediction Markets and Information Aggregation

Prediction markets aggregate dispersed information through the price mechanism (Hanson, 2003). Hanson advanced this framework by introducing combinatorial market designs and market scoring rules (MSR), demonstrating that automated market makers can efficiently maintain liquidity even across complex, joint probability spaces. However, these mechanisms rely on a critical assumption: that the event being bet upon is structurally definitive. Theoretical foundations establish that under certain conditions, market prices reflect the aggregate beliefs of participants, weighted by their willingness to bear risk. Empirical evidence demonstrates that prediction markets often outperform polls, expert forecasts, and statistical models in domains ranging from elections to box office revenues.

However, this literature largely assumes that contract specifications are unambiguous. This distinction maps to Ellsberg (1961) fundamental dichotomy between “risk” (where probabilities are known or estimable) and “ambiguity” (where the probability distribution itself is unknown). In a prediction market with poor syntax, traders face not only the risk of the event occurring, but the ambiguity of the resolution criteria. When participants disagree about what a contract means, rather than what outcome will occur, the information aggregation mechanism breaks down. Our work addresses this gap by quantifying linguistic ambiguity and measuring its effects on market function.

2.2 The Oracle Problem in Decentralized Finance

Decentralized prediction markets face the “oracle problem”: the challenge of importing real-world information onto blockchain systems in a trustworthy manner (Clark et al., 2014; Egberts, 2019). Duley et al. (2023) identify this as a systemic vulnerability in DeFi, noting that oracle manipulation attacks resulted in over \$403 million in losses in 2022 alone. Unlike centralized

platforms that rely on administrative resolution, decentralized markets use oracle mechanisms to adjudicate outcomes. UMA Protocol’s Optimistic Oracle operates on the principle that proposed data is assumed correct unless challenged within a defined liveness period; if disputed, the claim escalates to the Data Verification Mechanism (DVM), where UMA token holders vote using a Schelling point coordination game, voters are rewarded for aligning with the majority and penalized for voting with the minority or failing to vote (UMA Protocol, 2024).

Kleros employs a similar game-theoretic approach through its decentralized court system, where randomly selected jurors stake tokens and vote on disputes; jurors who vote with the eventual majority receive arbitration fees and tokens redistributed from incoherent voters, creating economic incentives for honest adjudication (Lesaege et al., 2019). Both systems rely on the assumption that truth serves as a natural Schelling point. Participants coordinate on honest answers because they expect others to do the same. These mechanisms typically involve token-holder voting, where disputes are resolved by majority consensus among staked participants. The oracle problem is often framed as a technical challenge of incentive design, ensuring that voters report truthfully. Our work highlights a complementary challenge: even with perfect incentives, ambiguous contract language can produce genuine disagreement among honest voters.

2.3 Contract Theory and Linguistic Ambiguity

The legal and economic literature on contracts has long recognized that natural language creates irreducible ambiguity. Hart & Moore (1988) developed the theory of incomplete contracts, demonstrating that parties cannot specify contingencies for all possible states of the world. Crucially, their model relies on *ex-post renegotiation* as a safety valve: when an unforeseen contingency arises, parties can voluntarily adjust terms to maximize mutual surplus. In blockchain-based prediction markets, this safety valve is structurally absent. Once a smart contract is deployed, its resolution logic is immutable; parties cannot “renegotiate” the terms when an ambiguity arises, turning what would be a solvable interpretation issue in traditional law into a zero-sum dispute.

Schwartz & Scott (2003) argue that commercial parties prefer a “formalist” (textualist) interpretation strategy to minimize variance and judicial intervention. Prediction markets take this preference to its logical extreme, often characterized as “code is law”, attempting to remove human discretion entirely to facilitate automation. However, this creates a paradox: while the execution layer (smart contracts) is perfectly formalist, the settlement layer (natural language definitions) remains subjective. Posner (2004) similarly frames interpretation as an economic trade-off: parties invest in drafting clarity only up to the point where the marginal cost equals the expected reduction in error and litigation costs. In prediction markets, where “litigation” (oracle disputes) freezes capital, this cost of error is uniquely high.

This theoretical framework directly informs our analytical approach. Since prediction markets lack the institutional safety valves of renegotiation (Hart & Moore) or equitable judicial discretion (Posner), the *ex-ante* clarity of the text becomes the sole determinant of market function. We therefore operationalize these concepts by treating “Semantic Risk” not merely as a linguistic feature, but as a quantifiable measure of *contract incompleteness*. By calculating the Semantic Risk Score (SRS), we empirically test the hypothesis that in the absence of legal recourse, market participants price this incompleteness as a distinct risk factor, withdrawing liquidity from contracts that fail to meet the formalist ideal.

2.4 Natural Language Processing in Finance

A growing literature applies natural language processing to financial documents. Loughran & McDonald (2011) demonstrated that standard dictionaries fail in financial contexts, necessitating domain-specific word lists. They found that words classified as “negative” in general English often

lack negative sentiment in 10-K filings. We extend this logic from vocabulary to semantics: prediction market clauses require context-aware interpretation that static word lists cannot provide.

Li (2008) established that syntactic complexity (or “readability”) correlates with poor firm performance, supporting the “management obfuscation hypothesis” where complex language is used to conceal adverse information. In prediction markets, we posit a similar mechanism: distinct from syntactic complexity, semantic ambiguity serves as a signal of low-quality market design. However, we distinguish complexity from ambiguity: a simple sentence like “Will Bitcoin rise?” is syntactically readable (low Fog Index) but semantically high-risk without a reference price or timeline.

Kearney & Liu (2014) survey methods for textual sentiment analysis, focusing on polarity (positive/negative) and tone. Our approach differs fundamentally: we measure interpretive ambiguity rather than emotional tone. We draw on recent work by Kuhn et al. (2023) on semantic uncertainty in language models, which distinguishes linguistic ambiguity from model uncertainty. Kuhn et al. demonstrate that models can be confident in their syntax while uncertain about the underlying meaning (“semantic entropy”). We adapt this theoretical distinction to the prediction market setting: our Semantic Risk Score captures the variance in *resolution outcomes* rather than the variance in token generation, effectively using the LLM as a proxy for the dispersed crowd of potential voters.

2.5 Ambiguity Aversion and Market Participation

Ellsberg (1961) established that decision-makers are averse to ambiguity, situations where the quality of information is too sparse to form a probability distribution, distinct from standard risk aversion. In equity markets, Bollerslev et al. (2009) demonstrate that the variance risk premium, the difference between implied and realized variance, captures a distinct form of risk that predicts returns through specific channels. Our semantic risk score plays an analogous role in prediction markets: it captures risk that is distinct from outcome uncertainty and manifests through identifiable market behavior. Kahneman & Tversky (1979) further documented the “certainty effect,” where decision-makers overweight outcomes perceived as certain. This bias helps explain our later finding that traders in ambiguous markets flock to perceived certainties (prices near 0 or 1). Our findings connect to this literature: traders appear to respond to semantic ambiguity by withdrawing from markets entirely, consistent with ambiguity aversion operating through participation rather than pricing.

Akerlof (1970) “Market for Lemons” provides our primary theoretical lens. In his model, asymmetric information about product quality leads to adverse selection and market collapse: sellers of high-quality goods withdraw because they cannot receive fair compensation when buyers cannot distinguish quality without incurring prohibitive verification costs, leaving only low-quality goods (“lemons”) in the market. Akerlof demonstrates that this dynamic can cause markets to unravel entirely, “the bad driving out the not-so-bad driving out the medium driving out the not-so-good driving out the good in such a sequence of events that no market exists at all”. We argue that semantic risk creates analogous dynamics in prediction markets: when contract clarity is difficult to assess, sophisticated traders exit because distinguishing well-specified contracts from poorly-specified ones requires costly analysis. This leaves a selected population of participants who either fail to perceive the ambiguity or are willing to bear the resolution risk, mirroring Akerlof’s adverse selection mechanism.

3 Theoretical Framework

3.1 Defining Semantic Risk

We define *semantic risk* as the probability that a prediction market contract will generate disagreement among resolvers due to linguistic ambiguity rather than factual uncertainty. Formally, consider a contract C with question text Q . Let $V = \{v_1, v_2, \dots, v_n\}$ be the votes of n oracle participants, where each $v_i \in \{0, 1\}$ represents a binary resolution decision.

The realized dispute variance is the Bernoulli variance of the vote distribution:

$$\text{Var}(V) = \hat{p}(1 - \hat{p}) \quad (1)$$

where \hat{p} is the proportion of affirmative votes. This variance ranges from 0 (unanimous agreement) to 0.25 (maximum disagreement at 50-50 split).

A clear contract produces low variance: voters agree on the interpretation, so disagreement reflects only idiosyncratic error. An ambiguous contract produces high variance: reasonable voters, applying different interpretive frameworks, reach different conclusions even when observing the same facts.

3.2 The Semantic Risk Score

We propose to estimate semantic risk using large language models. The key insight is that LLMs, trained on vast corpora of human text, have internalized patterns of linguistic ambiguity and interpretive disagreement. When an LLM produces high uncertainty about how a contract should resolve, this reflects genuine semantic ambiguity in the question text.

Our approach follows the few-shot learning paradigm (Brown et al., 2020). While Kuhn et al. (2023) measure “semantic entropy” over open-ended generations, we adapt their linguistic invariance principle to the binary prediction market setting by targeting the Bernoulli variance of the resolution outcome. Given a new question, the model predicts the expected variance:

$$\text{SRS}(Q) = \mathbb{E}[\text{Var}(V) \mid Q] \quad (2)$$

This predicted variance is the Semantic Risk Score. Higher SRS indicates greater linguistic ambiguity and higher probability of disputed resolution.

3.3 Hypotheses

We develop two hypotheses about how semantic risk affects market behavior:

Hypothesis 1 (Liquidity): Higher semantic risk leads to lower trading volume as sophisticated traders avoid contracts with ambiguous resolution criteria.

Hypothesis 2 (Convergence): Higher semantic risk leads to slower convergence, with prices remaining far from 0 or 1 due to interpretive disagreement.

As we will show, Hypothesis 1 is strongly supported, and Hypothesis 2 is supported but in the opposite direction predicted—a finding we explain through adverse selection dynamics.

4 Data and Methodology

4.1 Phase 1: UMA Protocol Validation Data

To validate our Semantic Risk Score, we require ground-truth data on actual dispute outcomes. We collect this from the UMA Protocol (UMA Protocol, 2024), a decentralized oracle system that resolves prediction market outcomes through token-holder voting.¹

¹Replication materials available at <https://github.com/rouzbeh-rs/semantic-risk-premium>

We query the UMA subgraph via The Graph API to obtain all resolved disputes from the protocol’s mainnet deployment. For each dispute, we extract the question text, vote distribution, and computed Bernoulli variance. After filtering for disputes with valid question text and sufficient vote participation, our validation sample comprises $N = 804$ resolved markets.

The distribution of actual variance is highly skewed: 87% of markets show low variance (< 0.05), indicating near-unanimous resolution, while a small tail exhibits substantial disagreement. This reflects the reality that most prediction market questions are reasonably clear; semantic risk is concentrated in a minority.

4.2 Phase 2: Polymarket Application Data

To test whether semantic risk affects market behavior, we collect data from Polymarket, the largest decentralized prediction market by volume. Using Polymarket’s Gamma API for market metadata and CLOB API for price history, we obtain resolved markets with complete trading data.

We apply several filters to ensure data quality: minimum volume of \$5,000, minimum duration of 3 days active trading, and valid price history with sufficient observations to calculate volatility metrics. Our final application sample comprises $N = 246$ markets with complete data for all variables.

For each market, we compute: (1) 24-hour realized volatility (standard deviation of returns in the 24-hour window ending 6 hours before resolution, to avoid the flat-price period after outcomes become known); (2) final price before resolution; (3) convergence gap, defined as $\min(p, 1 - p)$, measuring the distance of the final price from the nearest extreme; and (4) total trading volume.

4.3 SRS Estimation Methodology

We estimate SRS using Qwen3-8B, an open-source large language model with strong performance on reasoning tasks. We employ 4-bit quantization to enable efficient inference on consumer GPU hardware.

Following the few-shot learning paradigm (Brown et al., 2020; Sahoo et al., 2024), we construct a prompt containing 10 carefully selected examples from our UMA validation data. These examples span the full range of variance outcomes:

| Variance Category | Cate- gory | N | Exam- ples | Example Question |
|--------------------|---------------|---|---------------|---------------------------------------------------------------------|
| Zero (0.00) | | 5 | | MLB: New York Mets vs. Atlanta Braves 2023-04-29 |
| Low (0.02–0.03) | | 2 | | Will the highest temperature in London be 53°F or below on April 2? |
| Medium (0.08–0.10) | | 2 | | Will Trump launch a coin before the election? |
| High (0.21) | | 1 | | Farcaster unique users less than 15k on Feb 5? |

Table 1: Few-shot examples by variance category

The model is instructed to predict the Bernoulli variance (0.00–0.25) for each new question, considering factors such as definitional clarity, measurability, edge cases, and resolution criteria specificity. We use temperature 0.0 for consistent predictions.

4.4 Empirical Specifications

To validate SRS, we estimate the correlation between predicted and actual variance on the UMA holdout sample (excluding the 10 few-shot examples):

$$\text{Var}(V)_i = \alpha + \beta \cdot \text{SRS}_i + \varepsilon_i \quad (3)$$

To test hypotheses about market behavior, we estimate on the Polymarket sample:

$$\log(\text{Volume})_i = \alpha + \beta_1 \cdot \text{SRS}_i + \beta_2 \cdot \log(\text{Duration})_i + \varepsilon_i \quad (4)$$

$$\text{Convergence Gap}_i = \alpha + \beta_1 \cdot \text{SRS}_i + \beta_2 \cdot \log(\text{Volume})_i + \varepsilon_i \quad (5)$$

We winsorize continuous variables at the 1st and 99th percentiles to address outliers from thin trading or data anomalies.

5 Empirical Results

5.1 Validation: SRS Predicts Actual Dispute Variance

Table 2 presents the validation results. Our Semantic Risk Score correlates significantly with actual dispute variance in the UMA holdout sample ($n = 804$, excluding the 10 few-shot examples).

| Metric | Value |
|---------------------|-----------|
| Spearman ρ | 0.157 |
| p -value | < 0.001 |
| Pearson r | 0.157 |
| Mean Absolute Error | 0.0949 |
| N | 804 |

Table 2: SRS Validation Results (UMA Protocol)

While the correlation magnitude is modest, it is highly statistically significant and demonstrates that LLM predictions contain real signal about semantic ambiguity. The moderate effect size likely reflects our methodological choice to prioritize efficiency: we employ a 4-bit quantized version of Qwen3-8B, a relatively small open-source model, to enable rapid inference on consumer hardware. Larger frontier models (e.g., GPT-5, Claude) would likely yield stronger correlations, but at substantially higher computational cost. Our results establish a lower bound on achievable performance which means the signal is detectable even with a lightweight model. The mean absolute error of 0.0949 reflects the model’s tendency to over-predict variance for clear contracts (predicting ~ 0.14 when actual is 0.00) while performing well on genuinely ambiguous questions. Future work could explore whether larger models or fine-tuning on domain-specific dispute data improves prediction accuracy.

This validation establishes construct validity: our instrument measures what it claims to measure. Having validated the Semantic Risk Score on UMA data, we now apply it to Polymarket to test whether markets respond to this risk.

5.2 Hypothesis 1: Liquidity

Table 3 presents our central finding: semantic risk dramatically reduces trading volume.

The coefficient on SRS is large, negative, and highly significant ($\beta = -9.243$, $t = -4.41$, $p < 0.001$). Since volume is log-transformed, this implies that a 0.10 increase in SRS is associated with approximately a 60% reduction in trading volume ($\exp(-9.24 \times 0.10) \approx 0.40$).

| Variable | Coefficient | Std. Error | <i>t</i> -statistic | <i>p</i> -value |
|---------------|-------------|------------|---------------------|-----------------|
| Intercept | 9.277 | 0.387 | 23.79 | < 0.001 |
| SRS | −9.243 | 2.127 | −4.41 | < 0.001 |
| log(Duration) | 1.046 | 0.087 | 12.04 | < 0.001 |

Table 3: Liquidity Regression Results. $N = 244$. Dependent variable is $\log(\text{Volume})$.

This finding has a natural interpretation through the lens of [Akerlof \(1970\)](#). When traders cannot easily assess contract quality (clarity), they face adverse selection risk: entering an ambiguous market exposes them to potential disputes, frozen capital, and uncertain resolution. Rational traders respond by exiting, not by demanding a risk premium. The market doesn’t misprice semantic risk, it instead refuses to trade it.

5.3 Hypothesis 2: Convergence

Table 4 presents a surprising finding: high-SRS markets converge *faster* to extreme prices, not slower.

| Variable | Coefficient | Std. Error | <i>t</i> -statistic | <i>p</i> -value |
|-------------|-------------|------------|---------------------|-----------------|
| Intercept | 0.132 | 0.037 | 3.53 | 0.001 |
| SRS | −0.419 | 0.115 | −3.63 | < 0.001 |
| log(Volume) | −0.001 | 0.003 | −0.43 | 0.669 |

Table 4: Convergence Regression Results. $N = 244$. Dependent variable is convergence gap (distance from nearest extreme).

The coefficient on SRS is negative and highly significant ($\beta = -0.419$, $p < 0.001$). This means high-ambiguity markets end *closer* to \$0 or \$1 which means they exhibit greater apparent certainty, not less.

5.4 Interpretation: Selection-Driven False Consensus

The convergence finding initially appears paradoxical: how can ambiguous markets reach greater certainty? We propose a selection-based explanation grounded in adverse selection dynamics.

When a market has high semantic risk, sophisticated traders, those who recognize the ambiguity, rationally exit. They understand that disputed resolution could freeze their capital, and the expected value of participation is negative given the risk. This leaves a *selected* population of remaining participants who either: (a) fail to perceive the ambiguity; (b) hold unusually strong prior convictions; or (c) are speculating on short-term momentum. This aligns with findings by [Duley et al. \(2023\)](#), who caution that retail investors in DeFi lack the technical capacity to verify oracle integrity, making them disproportionately vulnerable to such opacity.

Crucially, this selected population exhibits belief homogeneity. Without informed counterparties to take the opposing view, prices converge rapidly toward an extreme. What looks like “certainty” in the price is actually the absence of dissent—false consensus driven by adverse selection rather than genuine information aggregation.

This interpretation unifies our two findings:

1. Strong liquidity effect: Exit manifests as collapsed volume.
2. Reversed convergence: Homogeneous remaining traders reach rapid (false) consensus.

This pattern mirrors [Akerlof \(1970\)](#) “Market for Lemons”: asymmetric information about quality (here, contract clarity) drives adverse selection and market failure. The prediction market analogue is that ambiguous contracts become lemon markets where only the least sophisticated participants remain. Akerlof drew an analogy to Gresham’s Law, “bad money drives out good”, noting that in his model, “bad cars drive out the good because they sell at the same price as good cars; similarly, bad money drives out good because the exchange rate is even”. In prediction markets, the parallel is that ambiguous contracts “drive out” clear ones in the sense that they dominate the attention of unsophisticated traders who cannot distinguish between them. The resulting price convergence in ambiguous markets is not informative consensus but rather the absence of informed dissent which is an outcome that may be mistaken for market efficiency but reflects selection rather than information aggregation.

6 Portfolio Optimization

6.1 The Capital Efficiency Problem

Traditional portfolio theory treats risk as return variance, though extensions have incorporated parameter and model uncertainty ([Garlappi et al., 2007](#)). In prediction markets, a different constraint binds: liquidity risk. When capital is deployed to an illiquid market, it is effectively frozen until contract resolution, which may be weeks or months away. During this period, the capital cannot be redeployed to other opportunities, incurring substantial opportunity costs.

Our findings establish that Semantic Risk Score is a strong predictor of market liquidity. This suggests incorporating SRS into portfolio construction to maximize capital efficiency.

6.2 Formal Framework

Let $M = \{1, 2, \dots, N\}$ be the universe of available prediction markets. For each market i , let:

- $\text{SRS}_i \in [0, 0.25]$ = Semantic Risk Score
- \hat{V}_i = Predicted trading volume given SRS
- w_i = Portfolio weight

From our liquidity regression, predicted log-volume follows:

$$\log(\hat{V}_i) = 9.277 - 9.243 \times \text{SRS}_i + 1.046 \times \log(\text{Duration}_i) \quad (6)$$

We formulate the portfolio selection problem as maximizing expected return subject to a semantic risk budget:

$$\max_{\{w_i\}} \sum_i w_i \cdot \mathbb{E}[R_i] \quad (7)$$

subject to:

$$\sum_i w_i = 1 \quad (\text{Budget constraint}) \quad (8)$$

$$w_i \geq 0 \quad (\text{No shorting}) \quad (9)$$

$$\sum_i w_i \cdot \text{SRS}_i \leq \bar{S} \quad (\text{Semantic risk budget}) \quad (10)$$

The semantic risk budget \bar{S} represents the investor’s tolerance for illiquidity and resolution uncertainty.

6.3 Clarity-Weighted Allocation

When expected returns are difficult to estimate (as in our setting without historical P&L data), we propose a simpler allocation rule that maximizes capital efficiency directly.

Define the *Clarity Score* as:

$$C_i = (1 - \text{SRS}_i)^\gamma \quad (11)$$

where $\gamma \geq 1$ is a risk aversion parameter. Higher γ concentrates allocation in the clearest markets.

The allocation weight for market i is:

$$w_i = \frac{C_i \cdot \mathbb{I}(\text{SRS}_i \leq \tau)}{\sum_j C_j \cdot \mathbb{I}(\text{SRS}_j \leq \tau)} \quad (12)$$

where τ is the clarity threshold and $\mathbb{I}(\cdot)$ is the indicator function. Markets with $\text{SRS} > \tau$ receive zero weight.

6.4 Determining the Optimal Threshold

Rather than selecting τ arbitrarily, we examine the liquidity relationship empirically. From our sensitivity analysis (Section 6.5), we observe that $\tau = 0.10$ provides an effective trade-off: it filters 59% of markets while more than doubling median volume. Lower thresholds (e.g., $\tau = 0.03$) yield higher median liquidity but reduce coverage to fewer than 60 markets, limiting diversification. Higher thresholds (e.g., $\tau = 0.15$) include too many ambiguous contracts, eroding the liquidity benefit. We adopt $\tau = 0.10$ as our recommended threshold, balancing portfolio diversification against semantic risk exposure.

6.5 Simulation Results

We compare two portfolios using our pilot dataset:

Naive Portfolio: Equal weight across all 246 markets

Clarity Portfolio: Clarity-weighted with $\tau = 0.10$ (filter high-SRS markets)

| Metric | Naive Portfolio | Clarity Portfolio | Improvement |
|------------------------|-----------------|-------------------|-------------|
| Markets Included | 246 | 100 | −59% |
| Mean SRS | 0.108 | 0.035 | −68% |
| Median Volume | \$59,567 | \$130,444 | +119% |
| Portfolio-Weighted SRS | 0.108 | 0.033 | −69% |

Table 5: Portfolio Simulation Results

The Clarity Portfolio dramatically improves liquidity characteristics. By filtering high-ambiguity contracts, median volume more than doubles. This means capital deployed to the Clarity Portfolio is substantially more likely to be recoverable and redeployable. This simulation demonstrates improved capital efficiency; testing return generation requires historical profit/loss data, which we leave for future work.

6.6 Practical Implementation

For practitioners, we distill our framework into a three-stage decision process:

Stage 1 (Semantic Screening): Compute SRS for each candidate market. Exclude markets where $\text{SRS} > 0.10$.

Stage 2 (Position Sizing): For eligible markets, allocate proportionally to $(1 - \text{SRS})^\gamma$.

Stage 3 (Monitoring): Re-score periodically; exit if SRS increases above threshold.

This framework can be implemented as an automated pre-trade check for trading systems or as a quality filter for market creation on prediction market platforms.

7 Conclusion

This paper introduces the Semantic Risk Score, a novel measure of linguistic ambiguity in prediction market contracts derived from large language model predictions. We validate this measure using ground-truth dispute data from the UMA Protocol and apply it to Polymarket to test whether semantic risk affects market behavior. Our findings reveal that semantic risk manifests as liquidity withdrawal: sophisticated traders exit ambiguous markets entirely, consistent with Akerlof (1970) adverse selection framework.

A counterintuitive finding emerges: high-ambiguity markets converge faster to extreme prices, exhibiting apparent certainty rather than confusion. We attribute this to selection effects: when sophisticated traders exit, the remaining homogeneous population rapidly reaches false consensus. This finding has important implications for interpreting prediction market signals; rapid convergence should not be mistaken for genuine information aggregation when it occurs in high-ambiguity contracts. For practitioners, we develop a portfolio optimization framework incorporating semantic risk as a liquidity factor. Filtering high-SRS contracts substantially improves capital efficiency, doubling median volume in our simulation. For platform operators, our results suggest that linguistic clarity should be a first-order concern in contract design, perhaps more important than topic selection or market-making incentives.

Future research could extend this work through cross-platform validation, larger samples enabling topic-level analysis, and trader-level data to directly verify the selection mechanism. Despite these limitations, our work establishes semantic risk as a quantifiable and economically significant factor in prediction markets. As decentralized finance continues to grow, and as prediction markets expand to cover increasingly complex domains, the tools we develop for measuring and managing linguistic ambiguity will become increasingly valuable.

References

- Akerlof, G. A. (1970). The market for “lemons”: Quality uncertainty and the market mechanism. *Quarterly Journal of Economics*, 84(3), 488–500.
- Bollerslev, T., Tauchen, G., & Zhou, H. (2009). Expected stock returns and variance risk premia. *Review of Financial Studies*, 22(11), 4463–4492.
- Brown, T., et al. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, 33, 1877–1901.
- Clark, J., et al. (2014). On the feasibility of decentralized derivatives. *Workshop on the Economics of Information Security (WEIS)*.
- Duley, C., et al. (2023). The oracle problem and the future of DeFi. *SSRN Working Paper*.
- Egberts, A. (2019). The oracle problem: An analysis of how blockchain oracles undermine the advantages of decentralized ledgers. *SSRN Working Paper*.
- Ellsberg, D. (1961). Risk, ambiguity, and the Savage axioms. *Quarterly Journal of Economics*, 75(4), 643–669.
- Garlappi, L., Uppal, R., & Wang, T. (2007). Portfolio selection with parameter and model uncertainty: A multi-prior approach. *Review of Financial Studies*, 20(1), 41–81.

- Hanson, R. (2003). Combinatorial information market design. *Information Systems Frontiers*, 5(1), 107–119.
- Hart, O., & Moore, J. (1988). Incomplete contracts and renegotiation. *Econometrica*, 56(4), 755–785.
- Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2), 263–291.
- Kearney, C., & Liu, S. (2014). Textual sentiment in finance: A survey of methods and models. *International Review of Financial Analysis*, 33, 171–185.
- Kuhn, L., Gal, Y., & Farquhar, S. (2023). Semantic uncertainty: Linguistic invariances for uncertainty estimation in natural language generation. *International Conference on Learning Representations (ICLR)*.
- Lesaege, C., Ast, F., & George, W. (2019). *Kleros Short Paper v1.0.7*. Retrieved from <https://kleros.io/whitepaper.pdf>
- Li, F. (2008). Annual report readability, current earnings, and earnings persistence. *Journal of Accounting and Economics*, 45(2-3), 221–247.
- Loughran, T., & McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *Journal of Finance*, 66(1), 35–65.
- Posner, R. A. (2004). The law and economics of contract interpretation. *Texas Law Review*, 83, 1581–1614.
- Sahoo, P., Singh, A. K., Srikanth, S., & Jain, A. (2024). A systematic survey of prompt engineering in large language models: Techniques and applications. *arXiv preprint arXiv:2402.07927*.
- Schwartz, A., & Scott, R. E. (2003). Contract theory and the limits of contract law. *Yale Law Journal*, 113, 541–619.
- UMA Protocol. (2024). *UMA Documentation*. Retrieved from <https://docs.uma.xyz/>