Supplementary Material

Decomposing Crowd Wisdom: Domain-Specific Calibration Dynamics in Prediction Markets

Nam Anh Le
Center for Applied Data Science and Artificial Intelligence,
College of Technology, National Economics University, Vietnam
me@namanhle.com

This document accompanies the main paper and provides (a) domain classification rules for both Kalshi and Polymarket, (b) cross-platform comparison tables, and (c) political subcategory slope analysis including a Simpson's paradox diagnostic. The full 216-cell calibration matrix is provided as a separate CSV file (`calibration_matrix_216.csv`).

Contents

1 Domain Classification Rules

Markets are classified into six analysis domains: Sports, Crypto, Politics, Finance, Weather, and Entertainment. Markets not matching any domain are classified as "Other" and excluded. The two exchanges require different classification strategies.

1.1 Kalshi: Prefix-Based Taxonomy

Kalshi assigns structured event tickers to every market (e.g., `KXNFLGAME-25-AFC-CHI-DET`). The leading alphanumeric prefix is extracted via:

```
regexp_extract(event_ticker, '^([A-Z0-9]+)', 1)
```

yielding, for example, `KXNFLGAME`. This prefix is looked up in a list of 571 pattern rules (`SUBCATEGORY_PATTERNS`), each mapping a prefix substring to a `(group, category, subcategory)` triple. Lookup is case-insensitive substring matching, with more specific patterns listed first. The implementation is:

```
def get_hierarchy(category: str) -> tuple:
    cat_upper = category.upper()
    for pattern, group, cat, subcat in SUBCATEGORY_PATTERNS:
        if pattern in cat_upper:
```

```
            return (group, cat, subcat)
    return ("Other", "Other", category)

def get_group(category: str) -> str:
    return get_hierarchy(category)[0]
```

Table 1 lists the six analysis domains with their principal ticker prefixes and descriptions.

Table 1: Kalshi domain composition: principal ticker prefixes.

Domain	Principal Prefixes	Description
Sports	`NFLGAME`, `NBAGAME`, `MLBGAME`, `NCAAFGAME`, `NCAAMBGAME`, `NHLGAME`, `WNBAGAME`, `UFCFIGHT`, `ATPMATCH`, `PGATOUR`, `EPLGAME`	Game outcomes, spreads, totals, player props, win totals
Crypto	`BTCD`, `BTCMAXY`, `ETHD`, `ETHMAXY`, `DOGE`, `SOL`, `XRP`, `SHIBA`, `COIN`	Daily/weekly price moves, token price thresholds
Politics	`PRES`, `SENATEAZ`, `HOUSEMOV`, `GOVPARTYVA`, `TRUMP`, `BIDEN`, `CABINET`, `MAYORNYCPARTY`	Elections, administration actions, policy, geopolitics
Finance	`FEDDECISION`, `INXU`, `NASDAQ100U`, `TNOTE`, `USDJPY`, `GAS`, `CPI`, `TARIFF`	Index daily moves, Fed decisions, macro indicators, commodities
Weather	`HIGHNY`, `RAINNYC`, `SNOW`, `TORNADO`, `HURCAT`, `ARCTICICE`	Temperature thresholds, precipitation, severe weather
Entertainment	`OSCAR`, `EMMY`, `GRAMMY`, `NETFLIX`, `BOX`, `SPOTIFY`	Awards shows, streaming metrics, box office

The full list of 571 prefix-to-domain mappings is in `src/classify.py` in the replication repository.

1.2 Polymarket: Regex Title Matching

Polymarket lacks structured ticker identifiers. Markets are classified by applying compiled regular expression patterns to the market title (question text). First match wins; patterns are ordered by specificity. Only four domains are used in the cross-platform comparison (Weather and Entertainment have negligible Polymarket coverage).

Sports pattern:

```
(?i)\b(NFL|NBA|MLB|NHL|UFC|MMA|boxing|tennis|golf|F1|
Formula 1|NASCAR|Super Bowl|World Series|Stanley Cup|
NCAA|March Madness|Premier League|Champions League|
La Liga|Serie A|Bundesliga|Ligue 1|Europa League|
World Cup|Copa America|MLS|WNBA|PGA|ATP|WTA|
Grand Slam|Wimbledon|US Open|Australian Open|French Open|
Ryder Cup|Olympics|Olympic|medal|playoff|postseason|
All[- ]Star|MVP|Cy Young|Heisman|touchdown|home run|
strikeout|rushing yards|rebound|assist|point spread|
moneyline|sack|interception|three-pointer|
batting average|ERA|free throw|penalty kick|
49ers|Packers|Chiefs|Eagles|Cowboys|Patriots|Dolphins|
Bills|Ravens|Bengals|Lakers|Celtics|Warriors|Bucks|
Nuggets|Knicks|76ers|Heat|Nets|Suns|Yankees|Dodgers|
Braves|Astros|Mets|Red Sox|Cubs|Phillies|Padres|
Maple Leafs|Bruins|Rangers|Lightning|Avalanche|Panthers|
Oilers|Arsenal|Liverpool|Manchester|Chelsea|Barcelona|
Real Madrid|Bayern|Juventus|
```

```
game \d|win.*series|win.*championship|win.*title)\b
```

Politics pattern:

```
(?i)\b(president|presidential|election|senate|congress|
governor|democrat|republican|GOP|Trump|Biden|Obama|
vote|ballot|primary|caucus|nominee|nomination|
Supreme Court|executive order|impeach|filibuster|
debt ceiling|shutdown|electoral college|swing state|
poll|approval rating|cabinet|attorney general|
secretary of state|Speaker|majority leader|veto|
legislation|midterm|runoff|recall|referendum|
RFK|DeSantis|Haley|Ramaswamy|Newsom|Pence|Kamala|
Harris|McConnell|Pelosi|AOC|Schumer|Vance|Vivek|
indictment|classified documents|NATO|Ukraine|Russia|
China.*Taiwan|Israel|Iran|sanction|TikTok.*ban|
government.*ban|Congress.*pass|federal|POTUS|
White House|inaugurat)\b
```

Crypto pattern:

```
(?i)\b(Bitcoin|BTC|Ethereum|ETH|crypto|Solana|SOL|
Dogecoin|DOGE|XRP|token|blockchain|DeFi|altcoin|
stablecoin|USDC|USDT|Tether|Binance|Coinbase|FTX|
SBF|halving|mining|NFT|airdrop|memecoin|Pepe|Shiba|
TVL|DEX|CEX|Bitcoin ETF|Ethereum ETF|spot ETF|
Layer 2|rollup|zkSync|Arbitrum|Polygon|Avalanche|
Cardano|ADA|Ripple|Litecoin|Polkadot|Chainlink|
Uniswap)\b
```

Finance pattern:

```
(?i)\b(S&P|S&P 500|NASDAQ|Dow Jones|Russell 2000|stock|
Fed |Federal Reserve|interest rate|rate cut|rate hike|
FOMC|CPI|inflation|GDP|recession|tariff|IPO|treasury|
yield curve|bond|unemployment|jobs report|
nonfarm payroll|non-farm payroll|PCE|PPI|
housing starts|retail sales|consumer confidence|
earnings|revenue|market cap|PE ratio|merger|
acquisition|bankruptcy|credit rating|oil price|
gold price|commodity|forex|quantitative|
balance sheet|FDIC|trade deficit|budget|fiscal|
monetary policy)\b
```

The classifier function applies patterns in order (Sports → Politics → Crypto → Finance) and returns "Other" if no pattern matches. This results in 42.5% of Polymarket markets classified as "Other," reflecting Polymarket's long tail of bespoke markets (celebrity events, tech product launches, meme markets).

2 Cross-Platform Comparison Tables

Table 2 presents the full cross-platform comparison of calibration slopes by domain and time horizon. Slopes are estimated from contract-weighted logistic regression with L_2 regularisation ($C = 10$). Cells with fewer than 200 trades are excluded.

Table 2: Cross-platform calibration slopes by domain and time horizon (Δ = Polymarket − Kalshi). Bins marked † are unreliable for Polymarket due to ∼3-hour timestamp noise from block-to-timestamp bucketing.

	Sports			Crypto			Politics		
Bin	Kalshi	Poly	Δ	Kalshi	Poly	Δ	Kalshi	Poly	Δ
0–1h†	1.101	1.028	−0.073	0.993	1.753	+0.760	1.341	0.621	−0.720
1–3h†	0.960	0.976	+0.016	1.013	1.390	+0.377	0.933	0.859	−0.073
3–6h	0.897	1.171	+0.275	1.065	0.862	−0.203	1.317	1.116	−0.201
6–12h	1.006	1.001	−0.005	1.007	1.044	+0.037	1.552	0.936	−0.616
12–24h	1.053	1.059	+0.006	1.006	0.996	−0.009	1.477	1.277	−0.200
24–48h	1.075	0.950	−0.126	1.209	1.123	−0.087	1.515	1.234	−0.281
2d–1w	1.037	1.097	+0.059	1.121	0.998	−0.123	1.833	1.982	+0.149
1w–1mo	1.240	0.974	−0.266	1.090	1.161	+0.071	1.833	1.681	−0.152
1mo+	1.740	1.322	−0.418	1.357	1.060	−0.298	1.730	1.086	−0.644
Mean (reliable)	1.150	1.082	−0.068	1.114	1.049	−0.065	1.637	1.313	−0.324

Table 3: Cross-platform calibration slopes for Finance (Δ = Polymarket − Kalshi). Polymarket Finance coverage is extremely thin (2,516 markets vs Kalshi's 38,058); these results are reported for completeness only.

Bin	Kalshi	Polymarket	Δ
0–1h†	0.957	0.914	−0.044
1–3h†	1.068	0.069	−0.999
3–6h	1.025	2.431	+1.406
6–12h	0.967	1.023	+0.057
12–24h	0.983	1.813	+0.830
24–48h	0.816	2.453	+1.636
2d–1w	1.065	1.676	+0.611
1w–1mo	1.420	1.756	+0.336
1mo+	1.200	2.113	+0.913

Note: The extreme Polymarket slopes (e.g., 2.431 at 3–6h, 2.453 at 24–48h) reflect very thin markets. Finance is excluded from main comparison figures and tables.

Table 4: Cross-platform whale effect bootstrap (Δ = mean slope for Large trades minus mean slope for Single trades, 10,000 bootstrap iterations).

Platform	Domain	Δ(L−S)	95% CI	Significant?
Kalshi (cell-level)	Politics	+0.531	[+0.288, +0.747]	Yes
Kalshi (market-clustered)	Politics	+0.589	[+0.132, +1.289]	Yes
Polymarket (cell-level)	Politics	+0.113	[−0.151, +0.395]	No
Polymarket (cell-level)	Sports	+0.006	[−0.199, +0.170]	No
Polymarket (cell-level)	Crypto	+0.092	[−0.158, +0.358]	No
Polymarket (cell-level)	Finance	+0.308	[−0.350, +0.910]	No

3 Political Subcategory Analysis

The aggregate Politics domain slope at the 1–3 hour bin ($\hat{b} = 0.933$) conceals extreme heterogeneity across subcategories. This section documents the subcategory composition and a Simpson's paradox diagnostic.

3.1 Subcategory Slope Ranges

Table 5 presents the 10 political subcategories sorted by their share of trades in the 1–3 hour bin. Slopes range from −0.136 (Biden Admin) to 7.219 (Senate), illustrating the compositional complexity hidden by the aggregate slope.

Table 5: Political subcategory slopes at the 1–3 hour horizon, sorted by trade share. The aggregate bin slope is 0.933.

Subcategory	Trades	% of Bin	Slope $\hat{b}$
Trump Admin	64,122	63.0	1.078
Presidential	14,439	14.2	1.040
Electoral College	7,560	7.4	1.807
Other Politics	5,932	5.8	2.169
Other Elections	3,865	3.8	0.175
Biden Admin	3,643	3.6	−0.136
House	1,006	1.0	0.567
Governor	758	0.7	1.438
Senate	376	0.4	7.219
NYC Mayor	126	0.1	5.621

3.2 Simpson's Paradox Diagnostic

Table 6 shows that the aggregate 1–3 hour slope (0.933) is below 1.0, but the two largest subcategories (Trump Admin at 1.078 and Presidential at 1.040, comprising 77.2% of trades) both have slopes above 1.0. This is a Simpson's paradox: subcategories with $\hat{b} > 1$ dominate the data, yet the aggregate slope is pulled below 1.0 by the nonlinear interaction of logistic regression with heterogeneous intercepts across subcategories. The leave-one-out column shows the aggregate slope when each subcategory is removed.

Table 6: Simpson's paradox diagnostic at the 1–3 hour horizon. Leave-one-out slope is the aggregate Politics slope after removing each subcategory.

Subcategory	Trades	% of Bin	Slope	Leave-One-Out Slope
Trump Admin	64,122	63.0	1.078	0.878
Presidential	14,439	14.2	1.040	0.944
Electoral College	7,560	7.4	1.807	0.689
Other Politics	5,932	5.8	2.169	0.891
Other Elections	3,865	3.8	0.175	1.009
Biden Admin	3,643	3.6	−0.136	1.183
House	1,006	1.0	0.567	0.934
Governor	758	0.7	1.438	0.929
Senate	376	0.4	7.219	0.931
NYC Mayor	126	0.1	—	0.933
Government	4	0.0	—	0.933
Cabinet	3	0.0	—	0.933

Note: Slopes for NYC Mayor, Government, and Cabinet are not estimated due to insufficient data at this horizon. Removing Biden Admin (3.6% of trades, slope −0.136) shifts the aggregate from 0.933 to 1.183, confirming that this single subcategory drives the below-unity aggregate.