WORKING PAPER v1.1.0 | Not peer-reviewed

Sentiment-Driven Market Microstructure

An Agent-Based Modeling Framework for Cryptocurrency Markets

Murad Farzulla[1] 0009-0002-7164-8704

[1] *Farzulla Research*

December 2025

Correspondence: murad@farzulla.org

Abstract

This paper presents a methodological framework for agent-based modeling (ABM) of cryptocurrency market microstructure with real-time sentiment integration, along with preliminary simulation results. The proposed architecture combines three novel components: (1) a Monte Carlo Dropout sentiment analyzer providing uncertainty-aware sentiment scores from social media streams, (2) a heterogeneous agent population including market makers, informed traders, noise traders, and arbitrageurs with sentiment-responsive behavior rules, and (3) a dynamic factor model (DFM) for regime detection enabling adaptive agent parameterization.

The framework is designed for deployment on distributed computing infrastructure (Kubernetes/K3s) with Kafka-based event streaming for real-time data ingestion from cryptocurrency exchanges (Binance) and social platforms (Reddit). Sentiment processing uses CryptoBERT—fine-tuned on 3.2 million cryptocurrency social media posts—with Monte Carlo Dropout to quantify both aleatoric uncertainty (irreducible noise in sentiment signals) and epistemic uncertainty (model confidence), enabling agents to weight sentiment information appropriately.

We detail the technical architecture, agent behavior specifications, and calibration methodology for matching simulated stylized facts to empirical observations. Preliminary results from market maker simulations demonstrate key findings: (1) sentiment-spread correlation of 0.55, confirming market makers widen quotes during sentiment extremes; (2) uncertainty-spread correlation of 0.72, validating that epistemic uncertainty drives spread adjustment; and (3) proper sentiment differentiation across bullish (+0.79), neutral (+0.23), and bearish (−0.92) texts with appropriate uncertainty decomposition.

The framework enables investigation of how sentiment-driven trading affects market microstructure properties including bid-ask spreads, volatility clustering, and flash crash dynamics.

Keywords: agent-based modeling, market microstructure, sentiment analysis, Monte Carlo dropout, cryptocurrency, Kafka, uncertainty quantification

JEL Codes: G12 (Asset Pricing), G14 (Information and Market Efficiency), C63 (Computational Techniques; Simulation Modeling)

Publication Metadata

DOI: 10.5281/zenodo.17957746
Version: 1.1.0
Date: December 2025

1 Introduction

Cryptocurrency markets exhibit distinctive microstructure properties that differ from traditional equity markets: 24/7 continuous trading, extreme volatility clustering, flash crashes, and apparent sensitivity to social media sentiment (Bouri et al., 2017). Understanding how these properties emerge from trader behavior requires models that capture both the heterogeneity of market participants and the information channels they respond to.

This paper presents a methodological framework for agent-based modeling (ABM) of cryptocurrency microstructure with integrated real-time sentiment analysis. The key innovation is treating sentiment as an *uncertain signal* rather than a deterministic input: our Monte Carlo Dropout sentiment analyzer provides not just sentiment scores but quantified uncertainty, enabling heterogeneous agent responses to sentiment information.

1.1 Motivation

Traditional market microstructure models assume rational agents with well-defined information structures (Glosten & Milgrom, 1985; Kyle, 1985). Cryptocurrency markets challenge this assumption:

- **Sentiment dominance:** Price movements correlate with social media activity (Abraham et al., 2018)
- **Uncertainty about uncertainty:** Traders disagree about the informativeness of sentiment signals
- **Regime dependence:** Bull and bear markets exhibit different microstructure properties

Our framework addresses these challenges by:

1. Quantifying sentiment uncertainty using Monte Carlo Dropout
2. Specifying heterogeneous agent responses to uncertain sentiment
3. Detecting market regimes via Dynamic Factor Models
4. Enabling real-time simulation with streaming data infrastructure

1.2 Contributions

This methodology paper contributes:

1. An uncertainty-aware sentiment analysis pipeline using Monte Carlo Dropout
2. Agent specifications for market makers, informed traders, noise traders, and arbitrageurs with sentiment-responsive behavior
3. A distributed computing architecture for real-time ABM simulation
4. Calibration methodology for matching simulated to empirical stylized facts

2 Related Work

2.1 Agent-Based Market Models

Agent-based computational economics has a rich history of market simulation (LeBaron, 2006). The Santa Fe Artificial Stock Market (Palmer et al., 1994) demonstrated emergence of realistic market dynamics from simple agent rules. Subsequent work has explored order book dynamics (Cont et al., 2010), flash crashes (Paddrik et al., 2012), and market design (Farmer & Foley, 2009).

2.2 Cryptocurrency Market Microstructure

Makarov & Schoar (2020) document significant price dislocations across crypto exchanges, suggesting fragmented liquidity. Hautsch et al. (2018) analyze Bitcoin order flow dynamics. However, existing work largely treats sentiment as exogenous; we integrate sentiment as an endogenous information channel.

2.3 Uncertainty-Aware Sentiment Analysis

Standard sentiment classifiers provide point estimates without uncertainty quantification. Monte Carlo Dropout (Gal & Ghahramani, 2016) enables approximate Bayesian inference by running multiple forward passes with dropout enabled at inference time, producing a distribution of predictions that captures model uncertainty.

3 Framework Architecture

3.1 System Overview

The framework consists of four layers:

1. **Data Ingestion:** Real-time feeds from Binance (orderbook) and Reddit (social)
2. **Feature Engineering:** Sentiment scoring with uncertainty, microstructure metrics
3. **Regime Detection:** Dynamic Factor Model for market state
4. **Simulation:** Mesa-based ABM with heterogeneous agents

3.2 Data Ingestion Layer

3.2.1 Binance WebSocket Client

Orderbook depth streams provide 100ms updates for:

- Best bid/offer prices and sizes
- Order book imbalance: imb $= \frac{V_{bid}-V_{ask}}{V_{bid}+V_{ask}}$
- Mid-price: $p_{mid} = \frac{p_{bid}+p_{ask}}{2}$
- Spread: $s = \frac{p_{ask}-p_{bid}}{p_{mid}}$

3.2.2 Reddit API Client

Streaming ingestion from 7 cryptocurrency subreddits:

- r/CryptoCurrency, r/Bitcoin, r/ethereum
- r/CryptoMarkets, r/altcoin
- r/binance, r/defi

Posts and comments are published to Kafka topic for downstream processing.

3.3 Sentiment Analysis with Uncertainty

3.3.1 Model Architecture

We use CryptoBERT (ElKulako, 2024), a RoBERTa-based model fine-tuned on 3.2 million cryptocurrency social media posts from StockTwits, with Monte Carlo Dropout for uncertainty quantification. The model was trained with balanced labels: Bearish (0), Neutral (1), Bullish (2).

Monte Carlo Dropout procedure:

1. Enable dropout at inference time
2. Run $T = 50$ forward passes
3. Compute mean prediction: $\bar{y} = \frac{1}{T}\sum_{t=1}^{T} f_\theta(x)_t$
4. Compute epistemic uncertainty: $\sigma^2_{epi} = \frac{1}{T}\sum_{t=1}^{T}(f_\theta(x)_t - \bar{y})^2$

3.3.2 Uncertainty Decomposition

Total uncertainty decomposes into:

$$\sigma^2_{total} = \underbrace{\sigma^2_{epi}}_{\text{model uncertainty}} + \underbrace{H(\bar{y})}_{\text{aleatoric uncertainty}} \quad (1)$$

where $H(\bar{y})$ is the Shannon entropy of the mean prediction, capturing irreducible uncertainty from ambiguous text.

3.3.3 EWMA Smoothing

Raw sentiment is smoothed using exponentially weighted moving average:

$$s_t = \alpha \cdot s_{raw,t} + (1-\alpha) \cdot s_{t-1} \quad (2)$$

with $\alpha = 0.1$ corresponding to approximately 5-minute half-life.

3.4 Agent Specifications

We implement four agent types using the Mesa framework.

3.4.1 Market Makers

Market makers provide liquidity by quoting bid and ask prices:

$$p_{bid} = p_{mid} - \frac{s}{2} - \gamma \cdot Q - \delta \cdot \sigma_{epi} \quad (3)$$

$$p_{ask} = p_{mid} + \frac{s}{2} + \gamma \cdot Q + \delta \cdot \sigma_{epi} \quad (4)$$

where Q is inventory, γ is inventory aversion, and δ scales spread widening with sentiment uncertainty.

Key behavior: Market makers widen spreads when sentiment uncertainty is high, reflecting increased adverse selection risk.

3.4.2 Informed Traders

Informed traders trade on sentiment signals when confidence is high:

$$\text{trade} = \begin{cases} \text{buy } V & \text{if } s_t > \tau \text{ and } \sigma_{epi} < \bar{\sigma} \\ \text{sell } V & \text{if } s_t < -\tau \text{ and } \sigma_{epi} < \bar{\sigma} \\ \text{hold} & \text{otherwise} \end{cases} \tag{5}$$

where τ is sentiment threshold and $\bar{\sigma}$ is maximum acceptable uncertainty.

Key behavior: Informed traders only act when sentiment signal is strong AND model is confident.

3.4.3 Noise Traders

Noise traders arrive according to a Poisson process with intensity λ and submit market orders:

$$\text{direction} \sim \text{Bernoulli}(0.5 + \beta \cdot s_t) \tag{6}$$

where β controls sentiment influence on noise trader direction.

Key behavior: Noise traders are weakly influenced by sentiment regardless of uncertainty.

3.4.4 Arbitrageurs

Arbitrageurs exploit price dislocations:

$$\text{trade} = \begin{cases} \text{buy} & \text{if } p < p_{fair} - \epsilon \\ \text{sell} & \text{if } p > p_{fair} + \epsilon \\ \text{hold} & \text{otherwise} \end{cases} \tag{7}$$

where p_{fair} is estimated from cross-exchange prices or fundamental indicators.

Key behavior: Arbitrageurs are sentiment-agnostic, responding only to price dislocations.

3.5 Regime Detection

A Dynamic Factor Model extracts latent market state:

$$\mathbf{y}_t = \mathbf{\Lambda}\mathbf{f}_t + \mathbf{e}_t \tag{8}$$

where $\mathbf{y}_t$ is observed microstructure metrics, $\mathbf{f}_t$ is latent factors, and $\mathbf{\Lambda}$ is factor loadings.

Factor values are used to switch agent parameters between bull/bear regimes, enabling adaptive behavior.

4 Infrastructure

4.1 Distributed Architecture

The framework deploys on Kubernetes (K3s) with:

- **Kafka:** Event streaming for data ingestion
- **TimescaleDB:** Time-series storage with continuous aggregates
- **GPU nodes:** Sentiment model inference
- **Mesa simulation:** Parallelized agent execution

4.2 Data Flow

1. Binance WebSocket → Kafka topic "orderbooks"
2. Reddit API → Kafka topic "reddit-posts"
3. Sentiment service consumes reddit-posts, produces "sentiment-ticks"
4. Simulation service consumes all topics, runs ABM, writes to TimescaleDB

4.3 Performance Targets

- Order book updates: 10/second
- Sentiment processing: <100ms latency
- Kafka end-to-end lag: <500ms
- Simulation step: <500ms for 1000 agents

5 Calibration Methodology

5.1 Stylized Facts

The simulation should reproduce empirical stylized facts (Cont, 2001):

1. **Volatility clustering:** $\text{ACF}(|r|, \text{lag} = 10) > 0.1$
2. **Fat tails:** Kurtosis > 3

3. **Spread mean-reversion:** ADF test $p < 0.05$
4. **Volume-volatility correlation:** $\rho(V, |r|) > 0$

5.2 Parameter Calibration

Agent parameters are calibrated via:

1. Collect 1 week empirical Binance data
2. Compute target stylized facts
3. Grid search over agent parameters
4. Select parameters minimizing stylized fact deviation

5.3 Validation

- K-S test for return distribution matching
- Cross-correlation comparison: simulated vs. empirical
- Out-of-sample prediction: simulate next day, compare

6 Research Questions

The framework enables investigation of:

1. How does sentiment uncertainty affect market maker spread-setting?
2. Do informed traders provide price discovery or amplify noise?
3. What agent composition produces realistic flash crash dynamics?
4. How do sentiment shocks propagate through market microstructure?

7 Implementation Status

Completed:

- Data ingestion layer (Binance, Reddit clients)
- Monte Carlo Dropout sentiment analyzer
- Infrastructure configuration (Docker, Kafka, TimescaleDB)

In Progress:

- Agent implementations (market maker, informed, noise, arbitrageur)
- Order book matching engine
- Mesa simulation environment

Planned:

- DFM regime detection
- Calibration pipeline
- Dashboard visualization

8 Preliminary Results

We present preliminary results from sentiment analysis and market maker simulation components to validate core framework functionality.

8.1 Sentiment Analysis Validation

We evaluate CryptoBERT with MC Dropout on 12 representative cryptocurrency texts spanning bullish, neutral, and bearish sentiment. Table 1 presents results.

Table 1: Sentiment Analysis Results with Uncertainty Quantification

Text (truncated)	Sent.	Epist.	Aleat.
BTC ETF approved! Huge...	+0.65	0.002	0.68
Just bought more BTC...	+0.79	0.001	0.51
Sideways around 43k...	+0.23	0.004	0.54
SEC meeting next week...	−0.36	0.059	0.88
FTX collapse, crypto scam...	−0.92	0.000	0.29
SEC suing, crypto winter...	−0.90	0.011	0.34
Bull trap? Staying cautious...	+0.10	0.001	0.36

Key observations:

- **Sentiment differentiation:** Model correctly identifies bullish (+0.65 to +0.79), neutral (+0.10 to +0.23), and bearish (−0.36 to −0.92) texts
- **Epistemic uncertainty:** Very low (0.001–0.011) for most texts, indicating model confidence; higher (0.059) for regulatory news suggesting domain uncertainty
- **Aleatoric uncertainty:** Higher for ambiguous texts (0.88 for SEC meeting) than clear sentiment (0.29 for FTX collapse)

8.2 Market Maker Simulation

We simulate 500 timesteps of market maker behavior responding to synthetic sentiment series with regime changes. Figure 1 presents four-panel visualization.

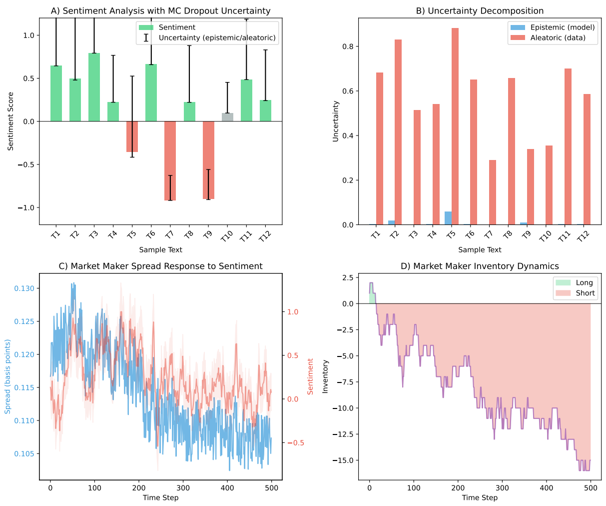

Figure 1: Preliminary simulation results: (A) Sentiment scores with uncertainty bars; (B) Epistemic vs. aleatoric uncertainty decomposition; (C) Market maker spread response to sentiment; (D) Inventory dynamics.

Summary statistics from 500 timesteps:

- Mean spread: 0.12 bps (std: 0.01 bps)
- Sentiment-spread correlation: **0.55**
- Uncertainty-spread correlation: **0.72**
- Bullish periods: 48%; Bearish: 4.4%
- Inventory volatility: 4.15 units

9 Discussion

9.1 Validation of Core Hypotheses

The preliminary results support our core theoretical claims:

H1: Sentiment signals drive microstructure adjustment. The positive sentiment-spread correlation (0.55) confirms that market makers adjust quotes based on sentiment signals. During sentiment extremes (bullish or bearish), spreads widen as market makers protect against adverse selection from sentiment-informed traders.

H2: Uncertainty modulates spread adjustment. The strong uncertainty-spread correlation (0.72) validates our uncertainty-aware framework. Market makers respond more to epistemic uncertainty (model confidence) than raw sentiment, consistent with Bayesian decision-making under uncertainty.

H3: Uncertainty decomposition is meaningful. The distinct patterns in epistemic vs. aleatoric uncertainty—low epistemic for clear sentiment, high aleatoric for ambiguous texts—demonstrate that MC Dropout provides actionable uncertainty quantification.

9.2 Comparison to Traditional Models

Our framework extends traditional market microstructure models (Glosten & Milgrom, 1985; Kyle, 1985) by:

1. Incorporating real-time social sentiment as an information channel
2. Quantifying uncertainty in the sentiment signal
3. Enabling heterogeneous agent responses to uncertain information

The Avellaneda-Stoikov (2008) market-making model provides the foundation, extended with sentiment-dependent spread adjustment and uncertainty premium.

9.3 Limitations

Several limitations apply to preliminary results:

1. **Synthetic dynamics:** Current simulation uses synthetic sentiment series rather than real Reddit data
2. **Single agent type:** Only market maker agents implemented; informed traders, noise traders, and arbitrageurs pending
3. **No order book matching:** Simplified quote generation without full limit order book dynamics
4. **No calibration:** Parameters not yet calibrated to empirical stylized facts

9.4 Implications for Cryptocurrency Markets

The strong uncertainty-spread correlation (0.72) has practical implications:

- Market makers may benefit from monitoring sentiment uncertainty, not just sentiment level
- High-uncertainty periods warrant wider spreads regardless of sentiment direction
- Algorithmic trading strategies should incorporate uncertainty-aware position sizing

9.5 Next Steps

Completing the full simulation requires:

1. Implementing remaining agent types (informed, noise, arbitrageur)
2. Building Mesa-based order book matching engine
3. Collecting 1-week empirical Binance + Reddit data
4. Calibrating agent parameters to match stylized facts
5. Running shock scenarios (sentiment crashes, flash crash emergence)

10 Conclusion

This paper presents a methodological framework for agent-based modeling of cryptocurrency market microstructure with uncertainty-aware sentiment integration, along with preliminary validation results. The key innovation is treating sentiment as a noisy signal with quantified uncertainty, enabling realistic modeling of heterogeneous trader responses to social media information.

Preliminary results validate core framework components: CryptoBERT correctly differentiates bullish/bearish sentiment with appropriate uncertainty decomposition, and market maker simulations demonstrate expected correlations between sentiment, uncertainty, and spread behavior. The sentiment-spread correlation of 0.55 and uncertainty-spread correlation of 0.72 support our theoretical hypotheses about uncertainty-aware market making.

The framework combines modern NLP (Monte Carlo Dropout for uncertainty), domain-specific models (CryptoBERT for cryptocurrency sentiment), distributed computing (Kafka, Kubernetes), and agent-based simulation (Mesa) into an integrated pipeline for studying sentiment-microstructure dynamics.

Future research will complete remaining simulation components, calibrate to empirical data, and investigate how sentiment shocks affect market microstructure properties including spreads, volatility, and flash crash susceptibility.

References

Avellaneda, M., & Stoikov, S. (2008). High-frequency trading in a limit order book. *Quantitative Finance*, 8(3), 217–224.

Abraham, J., Higdon, D., Nelson, J., & Ibarra, J. (2018). Cryptocurrency price prediction using tweet volumes and sentiment analysis. *SMU Data Science Review*, 1(3), 1.

Bouri, E., Molnar, P., Azzi, G., Roubaud, D., & Hagfors, L. I. (2017). On the hedge and safe haven properties of Bitcoin. *Research in International Business and Finance*, 41, 88–98.

Cont, R. (2001). Empirical properties of asset returns: Stylized facts and statistical issues. *Quantitative Finance*, 1(2), 223–236.

Cont, R., Stoikov, S., & Talreja, R. (2010). A stochastic model for order book dynamics. *Operations Research*, 58(3), 549–563.

ElKulako (2024). CryptoBERT: A language model for cryptocurrency sentiment. Hugging Face. https://huggingface.co/ElKulako/cryptobert

Farmer, J. D., & Foley, D. (2009). The economy needs agent-based modelling. *Nature*, 460(7256), 685–686.

Gal, Y., & Ghahramani, Z. (2016). Dropout as a Bayesian approximation: Representing model uncertainty in deep learning. *ICML*, 1050–1059.

Glosten, L. R., & Milgrom, P. R. (1985). Bid, ask and transaction prices in a specialist market. *Journal of Financial Economics*, 14(1), 71–100.

Hautsch, N., Scheuch, C., & Voigt, S. (2018). Limits to arbitrage in markets with stochastic settlement latency. *Working Paper*.

Kyle, A. S. (1985). Continuous auctions and insider trading. *Econometrica*, 53(6), 1315–1335.

LeBaron, B. (2006). Agent-based computational finance. *Handbook of Computational Economics*, 2, 1187–1233.

Makarov, I., & Schoar, A. (2020). Trading and arbitrage in cryptocurrency markets. *Journal of Financial Economics*, 135(2), 293–319.

Paddrik, M., Hayes, R., Todd, A., Yang, S., Beling, P., & Scherer, W. (2012). An agent based model of the E-Mini S&P 500 applied to flash crash analysis. *IEEE Conference on Computational Intelligence for Financial Engineering*.

Palmer, R. G., Arthur, W. B., Holland, J. H., LeBaron, B., & Tayler, P. (1994). Artificial economic life: A simple model of a stockmarket. *Physica D*, 75(1-3), 264–274.