1 Case Study: Factor Evolution Trajectory

This appendix presents a detailed case study of factor evolution in QuantaAlpha. We trace the complete trajectory of a representative factor—*Institutional_Momentum_Score_20D*—through the crossover phase, demonstrating how the evolutionary framework synthesizes complementary market hypotheses from parent trajectories.

QuantaAlpha's evolution process operates in three phases: (1) **Original** phase where initial hypotheses are generated, (2) **Mutation** phase where existing trajectories are perturbed to explore orthogonal strategies, and (3) **Crossover** phase where high-performing parent trajectories are combined to synthesize offspring with potentially superior predictive power. The following factor card illustrates a Round 8 Crossover operation.

1.1 Factor Identity

The factor card below presents the basic information of the evolved factor, including its unique identifiers, evolution lineage, and mathematical formulation.

🏷 Institutional_Momentum_Score_20D

Factor ID: c57cace576a95356
Trajectory ID: df5a496878f4
Evolution Round: Round 8
Evolution Phase: Crossover
Direction ID: 6

Factor Expression:

```
RANK(TS_CORR(DELTA($close, 1)/$close, DELTA($volume, 1)/$volume, 20) *
TS_MEAN(($close - $open)/$close, 5))
```

Mathematical Formulation:

$$\text{IMS}_{20D} = \text{RANK}\left(\rho_{20}\left(\frac{\Delta P}{P}, \frac{\Delta V}{V} \right) \times \overline{\left(\frac{C-O}{C} \right)}_5 \right)$$

where $\rho_{20}(\cdot, \cdot)$ denotes the 20-day rolling correlation, $\Delta P/P$ is the daily return, $\Delta V/V$ is the volume change ratio, and $\overline{(\cdot)}_5$ is the 5-day moving average.

Factor Interpretation:
This factor captures institutional-driven momentum by measuring two key signals: (1) the correlation between price returns and volume changes, which indicates coordinated institutional trading when positive; and (2) the average intraday return pattern, reflecting institutional activity that typically influences closing prices. The cross-sectional ranking ensures comparability across stocks.

1.2 Evolution Lineage

The crossover operation combines insights from two parent trajectories with complementary market hypotheses. Parent 1 focuses on identifying *fragile momentum* driven by retail speculation, while Parent 2 targets *sustainable momentum* supported by institutional activity. The LLM synthesizes these complementary perspectives into a unified framework.

⧗ Evolution Information

⑂ Parent Trajectories:

Parent 1: 1e6d57e38e89

Round: Round 7
Phase: Mutation
Rank IC: 0.0216
IC: 0.0059
IR: 1.297

Core Hypothesis:
When retail investors exhibit herd behavior and momentum chasing in stocks with high social media activity, but accompanied by declining institutional ownership and deteriorating fundamentals, the resulting price momentum is unsustainable and leads to mean reversion.
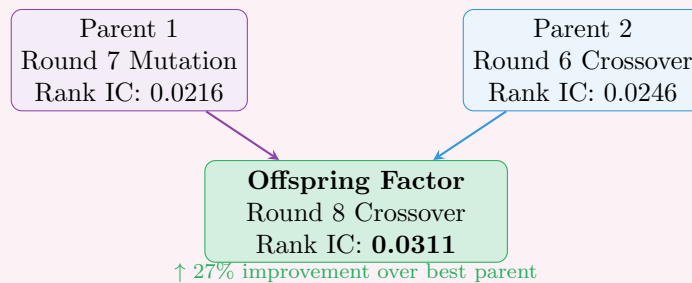
Parent 2: 47e0f0e55382

Round: Round 6
Phase: Crossover
Rank IC: 0.0246
IC: 0.0069
IR: 1.347

Core Hypothesis:
A regime-adaptive structural momentum factor combining institutional ownership-driven medium-term price trends with short-term microstructure regime validation, where coordinated accumulation/distribution patterns amplify momentum when confirmed by microstructure alignment.

⊞ Evolution Path Diagram:

```
Parent 1                     Parent 2
Round 7 Mutation             Round 6 Crossover
Rank IC: 0.0216              Rank IC: 0.0246
            \                /
             Offspring Factor
             Round 8 Crossover
             Rank IC: 0.0311
        ↑ 27% improvement over best parent
```

1.3 Synthesized Hypothesis

Through crossover, the LLM generates a new hypothesis that integrates the complementary insights from both parents, rather than simply averaging their factor expressions. This hypothesis-driven approach ensures that the offspring factor captures genuinely novel market dynamics.

> **💡 Hypothesis**
>
> **Core Hypothesis:**
> A regime-aware dual-source momentum factor that combines institutional-driven structural momentum (validated by healthy microstructure) and retail-driven speculative momentum (characterized by high attention and deteriorating fundamentals), dynamically weighted by market volatility: amplifying institutional signals in stable regimes and retail reversal signals in turbulent regimes, will generate superior predictive returns.
>
> | Component | Description |
> | --- | --- |
> | **Observation** | Parent strategies separately targeting institutional trends and retail herding show moderate predictive power (Rank IC ∼0.02–0.025), suggesting combined signals could capture complementary market dynamics. |
> | **Justification** | Sustainable price trends require institutional sponsorship and orderly trading, while retail-driven bubbles lack fundamental support and reverse under stress; a hybrid model exploiting both can enhance robustness across market regimes. |
> | **Domain Knowledge** | Institutional accumulation with strong price-volume correlation and low volatility indicates sustainable momentum; retail herding with declining institutional ownership and high volatility signals fragile momentum prone to reversal. |

1.4 Backtest Performance

After factor construction, QuantaAlpha automatically backtests the generated factors using the Qlib framework. The results below compare the offspring factor against both parent trajectories and the baseline, demonstrating the effectiveness of the crossover operation.

> **📈 Backtest Metrics**
>
> | Metric | Offspring Factor | Baseline | vs Baseline |
> | --- | --- | --- | --- |
> | **IC** | 0.0126 | 0.0058 | ↑ 117% |
> | **ICIR** | 0.0781 | – | – |
> | **Rank IC** | **0.0311** | 0.0220 | ↑ 41% |
> | **Rank ICIR** | 0.1932 | – | – |
> | **ARR (Excess)** | 7.80% | 5.20% | ↑ 50% |
> | **IR** | 0.963 | 0.973 | ↓ 1% |
> | **MDD (Excess)** | −11.37% | −7.30% | ↓ 56% |
>
> **Detailed Statistics:**
>
> | Metric | Value | Metric | Value |
> | --- | --- | --- | --- |
> | Daily Excess Return (w/o cost) | 0.0328% | Daily Excess Return (w/ cost) | 0.0128% |
> | Excess Return Std | 0.52% | Turnover (FFR) | 100% |
> | L2 Train Loss | 0.9936 | L2 Valid Loss | 0.9962 |

1.5 LLM Feedback and Iteration Guidance

After evaluating backtest results, the LLM provides structured feedback that guides subsequent evolution rounds. This feedback loop enables continuous improvement by learning from both successes

and failures.

> **💬 Feedback & Evaluation**
>
> **🔍 Observations:**
>
> The crossover operation produced mixed results compared to the baseline:
>
> - ✅ Higher annualized excess return (7.80% vs 5.20%)
>
> - ✅ Stronger IC (0.0126 vs 0.0058), indicating improved predictive power
>
> - ✅ Rank IC improved 27% over best parent (0.0311 vs 0.0246)
>
> - ❌ Larger maximum drawdown (-11.37% vs -7.30%)
>
> - ❌ Slightly lower information ratio (0.963 vs 0.973)
>
> **⚖️ Hypothesis Evaluation:**
>
> Results partially support the hypothesis. Improved annualized return and IC suggest that combining institutional and retail momentum signals has merit. However, deterioration in risk metrics indicates that without proper regime-adaptive weighting, the combined signals may amplify risks during turbulent periods. The full hypothesis requires all three components (institutional momentum, retail herding reversal, volatility-adaptive weighting) to work effectively.
>
> **🚩 Decision:** **REJECTED** for direct deployment
>
> **⚠️ Recommendations for Next Iteration:**
>
> 1. Use 20-day price-volume correlation as institutional momentum proxy
>
> 2. Use 5-day average intraday returns as retail attention proxy
>
> 3. Add volatility regime indicator (recent/historical volatility ratio) for dynamic weighting
>
> This feedback will inform the next mutation round, guiding the LLM to simplify the factor expression while preserving the core dual-source concept.

1.6 Key Takeaways

This case study demonstrates several key aspects of QuantaAlpha's evolutionary factor mining:

1. **Hypothesis-Driven Crossover**: The LLM synthesizes complementary market hypotheses rather than simply combining factor expressions, producing genuinely novel insights.

2. **Measurable Improvement**: The offspring achieved 27% Rank IC improvement over the best parent (0.0311 vs 0.0246), validating the crossover mechanism.

3. **Return-Risk Trade-off**: While predictive power improved, risk metrics deteriorated, highlighting the importance of balanced optimization.

4. **Structured Feedback**: Even rejected factors provide actionable guidance for subsequent iterations, enabling continuous improvement.

5. **Interpretability**: Throughout evolution, factors remain interpretable via natural language hypotheses and mathematical formulations.

This iterative process of hypothesis generation → factor construction → backtesting → feedback enables QuantaAlpha to systematically explore the factor space while maintaining interpretability and preventing overfitting.