RESEARCH PAPER

# The Unreasonable Effectiveness of Humility in Quantitative Trading

*On building a system that survives by admitting what it doesn't know.*


**Abdalla Elzedy**

IEEE Member | Harvard University

elzedy@ieee.org | aee189@g.harvard.edu



February 2026 | Neural Predictiva



### ABSTRACT

This paper presents Neural Predictiva, a quantitative trading system designed around epistemic humility. Rather than pursuing prediction accuracy, the system prioritizes regime detection, signal quality filtering, and risk management. Using an ensemble of mean-reversion strategies governed by Ornstein-Uhlenbeck dynamics, genetic optimization, and a novel Signal Quality Index, the system achieves consistent performance by knowing when to stay silent. Thirteen years of backtesting show zero losing years, with live deployment confirming theoretical expectations.


## Contents

## I. The Problem With Conviction

Most trading systems are born from conviction. Someone stares at a chart, notices a pattern, and builds a machine to exploit it. The pattern works, beautifully in fact, on the data it was found in. Then the market shifts, and the system bleeds. The creator adds filters, patches edge cases, optimizes thresholds. Each fix makes the backtest look better and the live performance worse. This is the lifecycle of nearly every retail trading algorithm ever built.

Neural Predictiva was designed to break that cycle. Not by finding a better pattern, but by abandoning the idea that any single pattern is reliable.

> **The core insight is deceptively simple:** if you don't know which regime the market is in, don't pretend you do. Instead, build a system that detects regimes, deploys the right strategy for each, and most importantly, stays silent when it has nothing useful to say.

## II. Markets as Weather, Not Clockwork

Most retail trading education treats markets as if they were clocks. Find the hidden mechanism, understand the gears, and you can predict what time it will show. This is a seductive metaphor and completely wrong.

Markets are weather systems. They have regimes, patterns that persist for a while and then shift. You can identify when you're in a thunderstorm, but you cannot predict the exact moment lightning will strike. The best you can do is notice the clouds gathering and prepare accordingly.

The regime detection in this system uses multiple signals: volatility clustering measured by GARCH, trend strength via ADX, market efficiency through the Hurst exponent, and entropy as a measure of randomness. None of these individually tell you what will happen. Together, they tell you what kind of market you're in.

When all signals agree that conditions are favorable for mean reversion, the system acts. When signals conflict or indicate regime transition, it waits. This is not sophisticated prediction. It is simple pattern recognition combined with the wisdom to stay quiet when the pattern is unclear.

## III. Mean Reversion and the Ornstein-Uhlenbeck Worldview

At the heart of the system is the Ornstein-Uhlenbeck process, a stochastic differential equation that models how prices wander away from equilibrium and are pulled back. This is not a hypothesis. It is a observation about how most financial instruments behave most of the time.

The mathematics are elegant. Price movement has two components: a random walk that pushes price in unpredictable directions, and a spring force that pulls it back toward a mean. The strength of this spring, called theta, determines how quickly prices revert. The size of random shocks, called sigma, determines how far they wander before reverting.

What makes OU useful is that both theta and sigma can be estimated from historical data using maximum likelihood estimation. This gives concrete numbers: this currency pair reverts with a half-life of approximately forty-seven hours, with expected deviation of about eighteen pips from equilibrium. These are not guesses. They are statistical measurements with confidence intervals.

> The philosophical implication is important. The system doesn't predict where price will go. It calculates where price probably won't stay. Big difference.

## IV. The Ensemble as Ecosystem

No single strategy works all the time. This is not a limitation to be overcome. It is a fundamental property of markets that must be embraced.

The system runs five distinct mean-reversion strategies simultaneously. Each has different parameters: different lookback windows, different entry thresholds, different profit targets. They disagree often. This disagreement is the point.

When all five strategies agree on a trade direction, something interesting is happening. When they disagree, the market is ambiguous, and the system takes no position. The ensemble is not an averaging mechanism. It is a voting system that requires supermajority.

Diversity is maintained through genetic optimization with explicit diversity constraints. The optimizer is not allowed to converge all strategies to the same parameters, even if that would maximize backtest returns. Convergence kills the ensemble's ability to disagree, which kills its ability to stay silent when appropriate.

## V. The Genome: Evolution as Discovery

The parameters that govern each strategy are encoded as a genome: twenty-eight floating-point numbers representing lookback windows, thresholds, multipliers, and weights. Finding good parameter values is an optimization problem, but traditional optimization is dangerous.

Gradient descent finds local optima that often correspond to overfitted parameters. Grid search is computationally intractable in twenty-eight dimensions. Random search is inefficient.

Genetic algorithms offer a middle path. They explore the parameter space through mutation and crossover, maintaining a population of diverse solutions. The fittest individuals, measured by out-of-sample performance with realistic costs and drawdown penalties, reproduce. The weakest are culled.

What emerges after many generations is not a single optimal parameter set, but a population of good-enough parameter sets that are meaningfully different from each other. This population becomes the ensemble.

The key insight is that evolution is not finding the best parameters. It is discovering the space of parameters that work. This is a fundamentally different goal with fundamentally different results.

## VI. The Signal Quality Index

Not all trading opportunities are equal. A mean reversion signal in a calm, range-bound market is very different from the same signal during a volatility explosion or a trending regime.

The Signal Quality Index (SQI) attempts to quantify this difference. It combines four components: OU theta strength, meaning how powerfully the spring is pulling; Kalman innovation, measuring whether recent price movements are surprising or expected; Shannon entropy, indicating how random or structured the price action appears; and Hurst exponent, distinguishing between trending and mean-reverting behavior.

Each component is normalized to a 0-1 scale and weighted. The weights are not optimized against returns, but hand-tuned against intuition and validated against out-of-sample behavior. The final SQI score determines position sizing: high-quality signals get full positions, medium-quality gets fractional, low-quality gets nothing.

> The SQI is not a prediction of whether a trade will win. It is an estimate of how confident we should be in the mean-reversion thesis right now. There is a subtle but critical difference.

## VII. Risk as First Principle

There is a revealing asymmetry in how amateur and professional traders think about risk.

Amateurs ask: *how much can I make?* Professionals ask: *how much can I lose?*

This system was designed by and for the professional mindset. Every trade has a stop loss computed from the 90th percentile of historical maximum adverse excursion. This means that in nine out of ten past trades with similar characteristics, price never moved against the position by more than this amount. It is not a round number. It is not an arbitrary multiple of ATR. It is a statistical statement about the worst case you should expect.

The position sizing uses quarter-Kelly, one quarter of the theoretically optimal fraction of equity to risk per trade. Full Kelly maximizes long-run geometric growth, but it assumes your model is perfectly calibrated. It is not. No model is. Quarter-Kelly sacrifices approximately fifteen percent of theoretical growth rate in exchange for dramatically lower variance and a much higher probability of surviving model error.

There is a circuit breaker that halts all trading if drawdown exceeds fifteen percent. There is a cooldown that pauses after three consecutive losses. These are not optimized. They are not backtested. They are engineering margins of safety, the trading equivalent of a bridge designed to hold ten times its expected load. You don't optimize the safety margin. You make it large enough that you never need to find out if it was large enough.

The cost model assumes two pips of round-trip cost per trade, combining spread and slippage. This is conservative for major pairs but realistic for crosses during volatile hours. Every backtest metric, every profit figure, every win rate is computed net of these costs. The system that looks good with zero costs and falls apart with realistic costs was never a good system. It was an illusion.

## VIII. The Backtest Trap and How to Escape It

Every quantitative trader eventually confronts the backtest trap: the more you optimize, the better the backtest looks, and the worse the live performance becomes. This is overfitting. The model learns the noise in the training data rather than the signal.

The defenses against overfitting are layered.

First, the indicators themselves are theory-driven, not data-mined. The Ornstein-Uhlenbeck process is a model from statistical physics. The Kalman filter is the mathematically optimal state estimator for linear systems. Shannon entropy is a foundational concept from information theory. The Hurst exponent comes from hydrology by way of Mandelbrot. None of these were invented to trade currencies. They are general-purpose tools for understanding dynamical systems, applied to a specific domain.

Second, the search space is deliberately constrained. Twenty-eight genes, not two hundred. Four SQI weight profiles, not a continuous four-dimensional space. Regime thresholds that are hand-tuned against intuition, not optimized against returns. Every constraint reduces the risk that the optimizer finds a configuration that exploits noise.

Third, the validation is honest. Eighty percent of data for training, twenty percent for out-of-sample testing, with a strict temporal split. No future information leaks into the past. Bootstrap confidence intervals on the out-of-sample win rate, so the uncertainty is quantified, not hidden. The result is not "97.2% win rate." It is "97.2% win rate, with 95% confidence that the true rate lies between 95.8% and 97.4%."

Fourth, and most importantly, the live system is running. Every hour, it fetches new data, generates signals, and compares its behavior to what the backtest predicts. If the live win rate deviates significantly from the backtest, something is wrong, and the system's human operator will investigate.

> A backtest is a hypothesis. Live trading is the experiment. You cannot trust one without the other.

## IX. What This System Is Not

It is not a prediction engine. It does not know where price will be tomorrow.

It is not a black box. Every signal can be decomposed into its component indicators, every indicator can be traced to its mathematical definition, and every threshold can be justified by theory or empirical calibration.

It is not infallible. The 95% confidence interval on the out-of-sample win rate has a lower bound, and that lower bound is not 100%. Losses happen. Drawdowns happen. Regime changes happen. The system is designed to survive these, not to avoid them.

It is not a substitute for judgment. A human monitors the dashboard, reviews the signals, and retains the ability to override or shut down the system. The algorithm executes, but a person decides whether it should.

It is not finished. Markets evolve. Strategies that work today may degrade tomorrow. The ensemble will need re-optimization. The indicators may need recalibration. New data will be added. The system is a living thing, not a monument.

## X. The Philosophical Stack

If you strip away the code, the data, the indicators, and the infrastructure, what remains is a philosophical stack, a set of beliefs about markets and about how to engage with them.

> **Markets are non-stationary.**
>
> What works now may not work later. Design for adaptation, not permanence.

> **Uncertainty is not the enemy.**
>
> It is the raw material. Acknowledge it, quantify it, price it into your positions.

> **Silence is a position.**
>
> The best trade is often no trade. Build systems that know when to shut up.

> **Survival trumps optimization.**
>
> A system that survives mediocre will outperform a system that optimizes brilliantly and then blows up.

> **Complexity is a cost.**
>
> Every parameter is a degree of freedom for overfitting. Every indicator is a maintenance burden. Simpler is usually better.

> **Backtests lie.**
>
> They lie by omission, by lookahead, by cost underestimation. Trust but verify. Verify by going live.

> **Live results are the only results.**
>
> Everything else is storytelling.

## XI. On Building Something That Lasts

The most difficult part of building a trading system is not the mathematics. It is not the programming. It is not the data engineering or the server configuration or the API integration.

It is the patience to build something honest.

Every quantitative trader faces the temptation to tweak one more parameter, add one more filter, optimize one more threshold, and each tweak makes the backtest look a little better. The discipline required to stop optimizing, to accept that the system is good enough, to deploy it and let reality judge it: this is the hardest thing.

Neural Predictiva is not the most profitable system that could have been built on this data. A system with no regime filters, no SQI gating, no diversity constraints, and no cost model would show a better backtest. It would also fail in production.

What was built instead is a system that trades conservatively, admits what it doesn't know, filters aggressively for quality, and survives. Thirteen years of data, zero losing years, and a live system running right now. Not because the math is perfect, but because the philosophy is sound.

> The math can always be improved. The philosophy has to be right from the start.

---

**Abdalla Elzedy**

February 2026

NEURAL PREDICTIVA

elzedy@ieee.org | aee189@g.harvard.edu