



A Human In The Loop Alpha Generation

Rose Rahimi

Adviser: Valerie Barr & Emanuele Citera

Computational Science Program, Bard College, May 2026



Motivation

Distributional generative models can produce profitable trading signals from noisy markets, but performance is sensitive to architecture, training dynamics, and evaluation choices.

FinGAN models the *distribution* of future returns rather than point forecasts, enabling confidence-aware position sizing. Its 10 step return window limits the historical context the generator can draw on. Naively lengthening the window does not work: the recurrent layer forgets early signal by the final timestep.

Key Contributions

- **TF2 Re-implementation:** of Vuletic's PyTorch FinGAN, modular and scriptable.
- **FinTimeGAN.** Embedder & Recovery extension running the adversarial game in a learned latent space.
- **Branched Loss Protocol:** Each loss configuration is trained independently from a shared initialization.
- **Return Representation Study:** Compares relative excess and beta adjusted excess returns.
- **Three Gate Evaluation:** Selection, statistical significance, distributional sanity.

Data & Universe

The 28 individual ticker universe spans nine GICS sectors. Targets are next half day relative excess returns. Train / validation / test split is chronological 80 / 10 / 10. Test window covers 1,098 trading days, 2019-10-30 to 2021-12-31.

ETF Tickers	Sector Name	Stocks (Ticker, Company Name)
XLY	Consumer Discretionary	(AMZN, Amazon.com), (CCL, Carnival), (EBAY, eBay), (HD, Home Depot), (NKE, Nike)
XLP	Consumer Staples	(CL, Colgate-Palmolive), (EL, Estée Lauder), (KO, Coca-Cola), (PEP, PepsiCo), (SYY, Sysco), (TSN, Tyson Foods)
XLE	Energy	(APA, APA Corp.), (OXY, Occidental Petroleum)
XLF	Financial	(BLK, BlackRock), (GS, Goldman Sachs), (TROW, T. Rowe Price), (WFC, Wells Fargo)
XLV	Health Care	(CERN, Cerner), (HUM, Humana), (PFE, Pfizer)
XLI	Industrials	(FDX, FedEx), (GD, General Dynamics)
XLK	Information Technology	(IBM, IBM), (TSLA, Tesla)
XLB	Materials	(ECL, Ecolab), (IP, International Paper)
XLU	Utilities	(DTE, DTE Energy), (WEC, WEC Energy)

Table 0.1: The 28 stock universe organized by sector ETF benchmark.

Loss Functions

Four economic driven loss terms beyond standard BCE. Each measures a quantity a trader cares about. Starred quantities are smooth, differentiable proxies of their financial counterparts.

Term	Measures
PnL*	Smoothed profit and loss; $\tanh(x/2)$ replaces sgn
MSE	Magnitude calibration; anchors forecast scale
SR*	Smoothed Sharpe ratio of the strategy PnL series
STD	Standard deviation of strategy PnL series

Ten combinations train independently in Phase 2b: BCE alone, PnL*, PnL*+MSE, PnL*+MSE+STD, PnL*+MSE+SR*, PnL*+SR*, PnL*+STD, SR*, SR*+MSE, MSE.

Models

Five forecasting models evaluated under a single protocol on the same 549 day test window.

- **ARIMA(2,0,2).** Order chosen by AIC over $\{(1,0), (1,1), (2,0), (2,1), (2,2)\}$ after ADF stationarity testing, with rolling one step ahead forecasts at fixed coefficients.
- **Ridge AR(10).** Regularized linear autoregression on the 10 step condition window.
- **LSTM-Fin.** Recurrent point forecast; six economics driven loss variants.
- **Fin-GAN.** Vuletic et al.'s reference cGAN; adversarial game in *return space*.
- **FinTimeGAN.** Adds a TimeGAN style Embedder & Recovery pair. Generator and discriminator operate in a learned latent space. Two context modes: *compressed* (final latent vector h_{last}) and *attention* (full latent sequence via multi-head attention).

Model	Distributional	Recurrent	Latent space
ARIMA(2,0,2)	—	—	—
Ridge AR(10)	—	—	—
LSTM-Fin	—	✓	—
FinGAN	✓	✓	—
FinTimeGAN	✓	✓	✓

Trading Rules

Trading rules. A generator produces $B = 1,000$ Monte Carlo samples per time window. Two ways to convert that distribution to a position:

Rule	Position	Applies to
Hard signal	$h_t = \text{sgn}(x_t) \in \{-1, 0, +1\}$	All models
Weighted signal	$w_t = p_t^{(0)} - p_t^{(1)} \in [-1, +1]$	Distributional only

Three Gate Evaluation

A model passes if and only if it clears all three gates.

Selection	Significance	Sanity
Validation best loss combination per (ticker, seed) by validation weighted signal Sharpe	200 permutation null on $\text{sign}(\text{PnL}_t)$ at the portfolio level; reject at $p < 0.05$.	Four collapse flags catch degenerate generators (narrow distributions, sign-degenerate output).

Descriptive diagnostics (reported alongside every result, not used as gates):

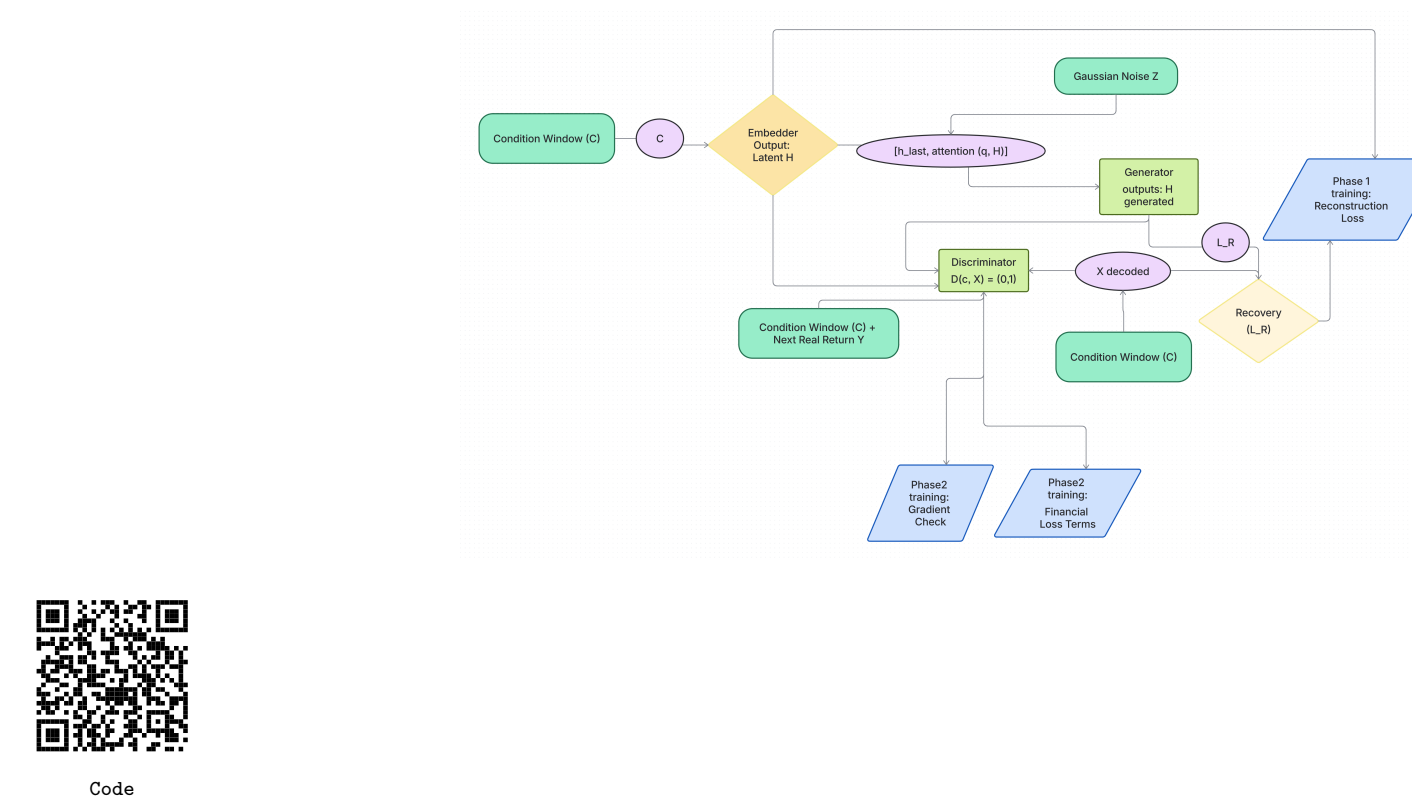
Metric	What it tests
Constant sign Sharpe SR_{const}	Rules out market drift; if model $\approx$ constant long, no skill
Hit rate	Directional accuracy; chance is 0.50; threshold 0.53
MAE / RMSE	Magnitude calibration; catches tanh saturation artifacts
Forecast target correlation ρ	Tracking signal vs. drift; indifferent to position size
Per ticker permutation p	Tiered: strong (< 0.05), suggestive (< 0.20), inconclusive

Training Phases

Phase 1 Pretraining	Phase 2a Calibration	Phase 2b Adversarial
Train Embedder E and Recovery R on reconstruction loss.	Record gradient norms of each loss term over 25 epochs.	Train G, D on each of 10 loss combinations independently.
Output: latent space $H = E(c)$	Output: scaling coefficients $\alpha, \beta, \gamma, \delta$	Output: per-combo snapshots $\{(d_k, \hat{d}_k)\}$
E, R frozen for Phase 2.	Freeze checkpoint θ^* .	Each branch restores θ^* , reseedes RNG.

FinTimeGAN Architecture

Two stage design: Phase 1 pre-trains the Embedder & Recovery pair as an autoencoder on real returns. Phase 2 freezes both networks and runs the adversarial game in latent space.

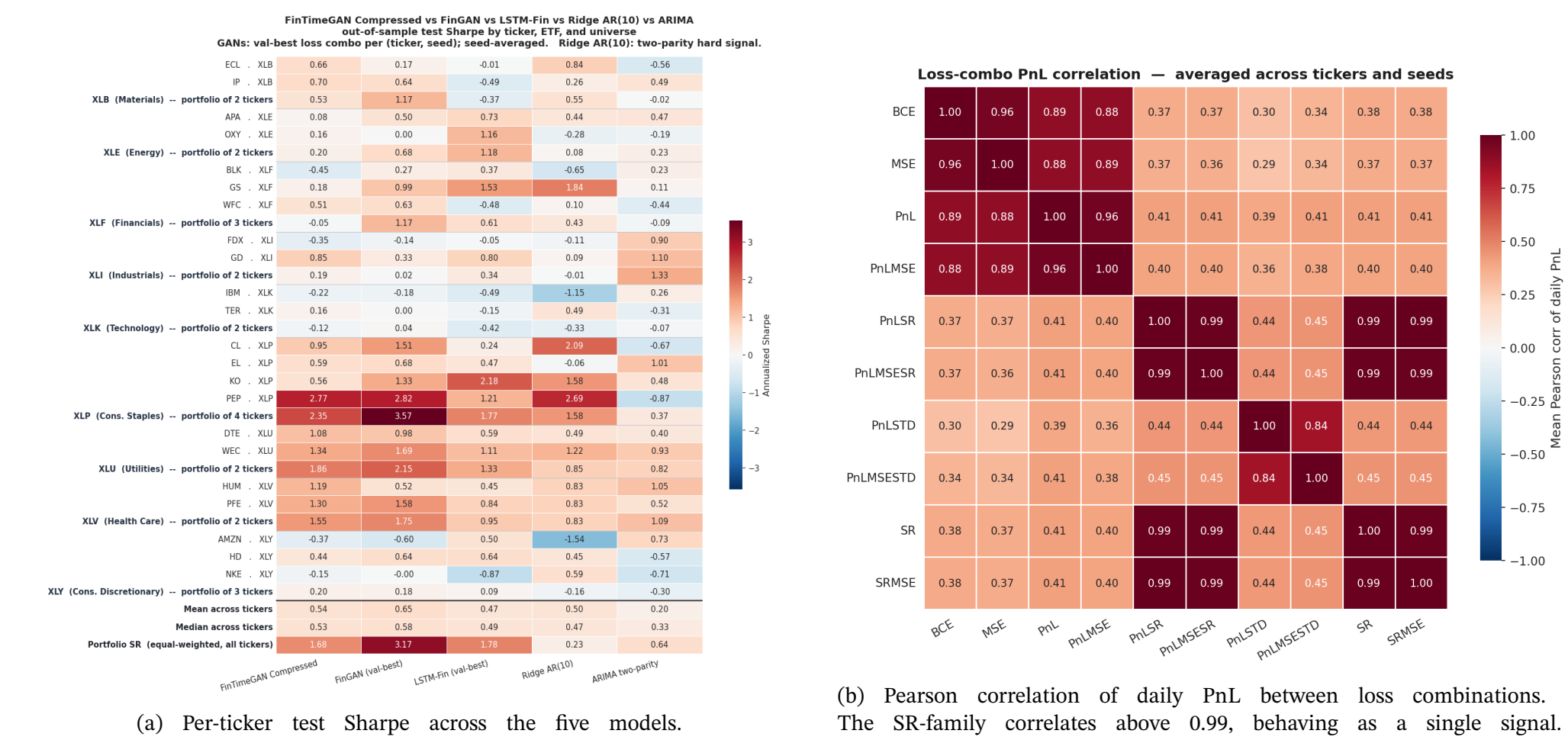


Code

Results & Findings

Model	Universe	Portfolio Sharpe	p-value
ARIMA(2,0,2)	28 stocks	+0.642	
Ridge AR(10)	28 stocks	+0.227	
LSTM-Fin	28 stocks	+1.370	$< 10^{-4}$
FinTimeGAN (compressed)	27 stocks	+1.083	0.040
FinGAN	28 stocks	+2.022	$< 10^{-4}$
FinGAN	22 stocks + 9 ETFs	+3.17	$< 10^{-4}$

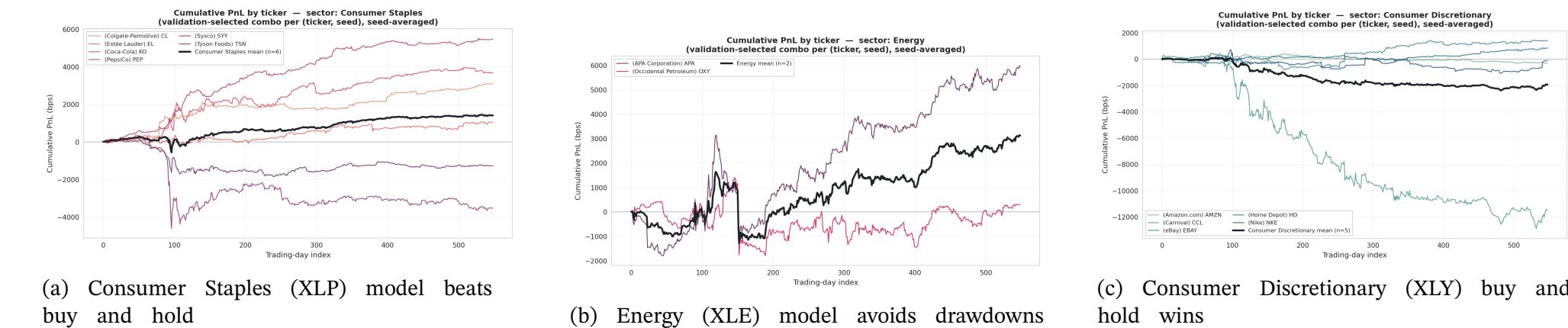
Table 0.2: Portfolio weighted signal Sharpe across all five models. Significance from a 200 permutation null on $\text{sign}(\text{PnL}_t)$.



(a) Per-ticker test Sharpe across the five models. (b) Pearson correlation of daily PnL between loss combinations. The SR-family correlates above 0.99, behaving as a single signal.

FinGAN's Cumulative PnL

Sector level cumulative PnL on 22 stocks + 9 ETFs universe, where FinGAN reaches portfolio Sharpe +3.17.



Three patterns in one universe. FinGAN's value takes two forms in defensive sectors and drawdown avoidance in volatile cyclicals; growth tilted sectors are where the model loses to passive long exposure. The cross sectional pattern is consistent across all five models.

References

- [1] *Multi-asset financial markets: mathematical modelling and data-driven approaches*, University of Oxford, 2025.
- [2] *Deep Learning*, section 5.5, pp. 131-134, The MIT Press, 2016.