7 Steps to Systematic Investing

Jeff Ryan | Justin M Shea

Data Collection and Preparation

Identify Data Sources:

- Determine which market, fundamental, or alternative datasets are needed (e.g., intraday prices, macroeconomic indicators, Form 10k financials, etc).

Gather and Standardize:

- Acquire data from APIs, databases (Bloomberg, Refinitiv, quantKiosk, Quandl, CoinGecko), ensuring consistent timestamps and reference IDs.

Clean and Validate:

- Check for missing values, outliers, or stale data that could skew outcomes. Apply standard cleaning procedures (e.g., interpolation for time-series gaps). Feature Engineering:
- Generate features suited to the intended model type (e.g., technical indicators, factor exposures, and computing returns.

Exploratory Data Analysis (EDA)

Data Distribution & Summary Stats:

- Examine mean, variance, skewness, and kurtosis of inputs to confirm alignment with the model's assumptions.

Correlation Analysis:

- Identify overlapping signals or redundant features that might impact performance.

Visualization:

- Plot sample outputs or signals on historical data to see if they capture important turning points, trends, or market regimes.

Hypothesis Formulation:

- Based on observations, hypothesize how the model might capture market inefficiencies (e.g., patterns in volatility or cyclical behaviors).

Model Selection and Development

Understand the Model’s Architecture:

- If it’s a machine learning model, clarify whether it’s a neural network, ensemble method, or specialized time-series algorithm.
- If it’s a statistical model, identify key assumptions (e.g., stationarity, normality of residuals).

Implementation Details:

- Adapt raw data or engineered features into the format required by the modeling framework or packages (e.g., sequences for an LSTM, tabular for an XGBoost). See model documentation.
- Hyperparameter Tuning:
- Employ grid search or Bayesian optimization, using cross-validation for robust parameter selection.

Domain-Specific Constraints:

- Ensure no lookahead bias and minimal data snooping related to your domain

Backtesting, Simulation, and Validation

Historical Simulation:

- Evaluate performance metrics (e.g., Sortino ratio, maximum drawdown) on past data segments.

Walk-Forward Analysis:

- Retrain periodically on rolling windows to mimic real-world conditions and test adaptability.

Out-of-Sample Testing:

- Reserve distinct data (e.g., the last 6–12 months) to verify how well the model generalizes to unseen market conditions.

Performance Metrics: Returns, CAGR, P\&L distribution, etc relevant to strategy/asset class

Risk: standard deviation, Conditional Value at Risk (cVaR, VaR), Drawdowns, etc

Stability: Statistical tests for overfitting

Risk Management and Stress Testing

Parameter Sensitivity:

- Test how changes in key parameters affect the model's risk exposures.

Scenario Testing:

- Evaluate performance under different market regimes (e.g., financial crises, high-volatility periods).

Stress Test Extremes:

- Shock critical variables to test model ouput

Implementation and Deployment

Coding & Integration:

- Convert your prototype into production-ready code with robust error handling and logging.

Infrastructure Setup:

- Deploy on cloud platforms (AWS, GCP, Azure) or on-premise HPC, ensuring sufficient compute resources for the model’s needs.

Execution Mechanisms:

- Connect the model’s signals to your broker or in-house trading engine.
- Monitor transaction costs, slippage, and latency to preserve realistic performance.

Compliance & Documentation:

- Thoroughly document the model’s logic, assumptions, and performance.
- Ensure alignment with regulatory standards (CFTC, MiFID II, SEC, etc.)

Monitoring and Maintenance

Live Performance Monitoring:

- Compare real-time signals and executed trades against backtest baselines.
- Track continuous risk measures (leverage, exposure, margin).

Model Drift Detection:

- Observe data distribution shifts or market condition changes that may degrade performance.
- Measure half-life of strategies over time and performance degradation

Ongoing Retraining:

- Periodically retrain on the latest data to incorporate new market behavior.

Reporting and Review:

- Produce daily, weekly, or monthly performance summaries.
- Conduct regular strategy reviews to decide whether to continue, adjust, or retire the model.

Thank you for engaging