2024 Edition

Automated Trading System

Applies cointegration method for statistical arbitrage in the Brazilian stock market, optimized through econometrics, data mining and machine learning

Section 01

Presentation

strategic business plan

The trading system exists to exploit market inefficiencies through disciplined pair trading strategies, delivering consistent, risk-adjusted returns that exceed the performance of the collateral providing margin. It aims to be a high-performance automated trading system, generating stable returns by leveraging advanced statistical models and data-driven insights.

The project focus on achieving superior returns through precise pair trading strategies, backed by rigorous monitoring and robust statistical analysis. Precision, risk management, system reliability, and a commitment to continuous improvement using cutting-edge techniques in statistical modeling and market insights.

It reaches goals by consistently backtesting, improving the algorithms, integrating machine learning for better pattern recognition, and ensuring seamless execution on the bot platform.

Backtesting is conducted over 10 years of daily data. The strategy is evaluated based on Sharpe Ratio, Profit, and Drawdown.

Results will be presented through periodic reports summarizing key performance metrics, including detailed trade outcomes and deviations from expected results.

Success will be defined by consistent performance X% above the collateral benchmark, with controlled risk (Sharpe Ratio > 1.5, Drawdown < 20%).

What is being used and further steps:

Framework: Statistical arbitrage approach using cointegration variables and z-scores for signal generation.

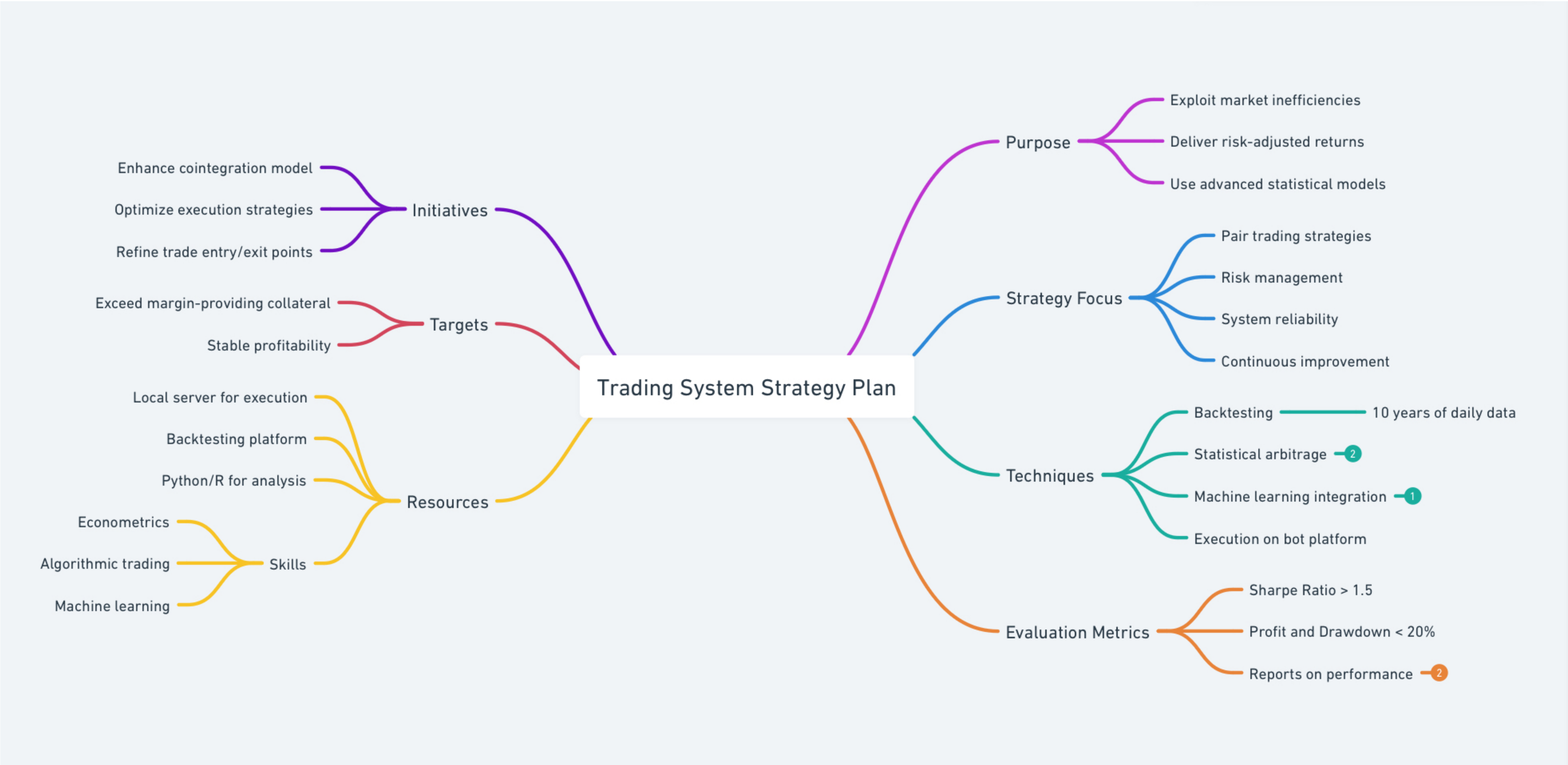
Resources: Local server for execution, backtesting platform, and tools like Python/R for analysis.

Skills: Proficiency in econometrics, algorithmic trading, machine learning, and data analysis in Python and R.

Targets: Achieve stable profitability, exceeding the performance of margin-providing collateral.

Initiatives: Continuously enhance the cointegration model, optimize execution strategies, and integrate machine learning to refine trade entry/exit points.

mind map



A **mind map** is a visual tool used to organize information, thoughts, or ideas around a central concept, facilitating brainstorming, planning, and understanding complex topics. It typically starts with a central idea or keyword, branching out into related subtopics, each of which may have further sub-branches. These interconnected nodes form a radial structure that can resemble the branches of a tree or a network. Mind mapping is commonly used in education, business, creative thinking, and personal productivity. By using mind maps, individuals and teams can simplify complex topics, discover new ideas, and streamline workflows, leading to better outcomes and increased efficiency.

statistical model

At the core of econometrics, the **statistical model** serves as the fundamental framework for analyzing relationships between time-series. On the right side you can see the cartesian plot between a dependent variable **price.y** and a independent variable **price.x** and a **linear regression** establishes the relationship. This model allows economists and financial analysts to identify underlying patterns, test hypotheses and forecast future trends.

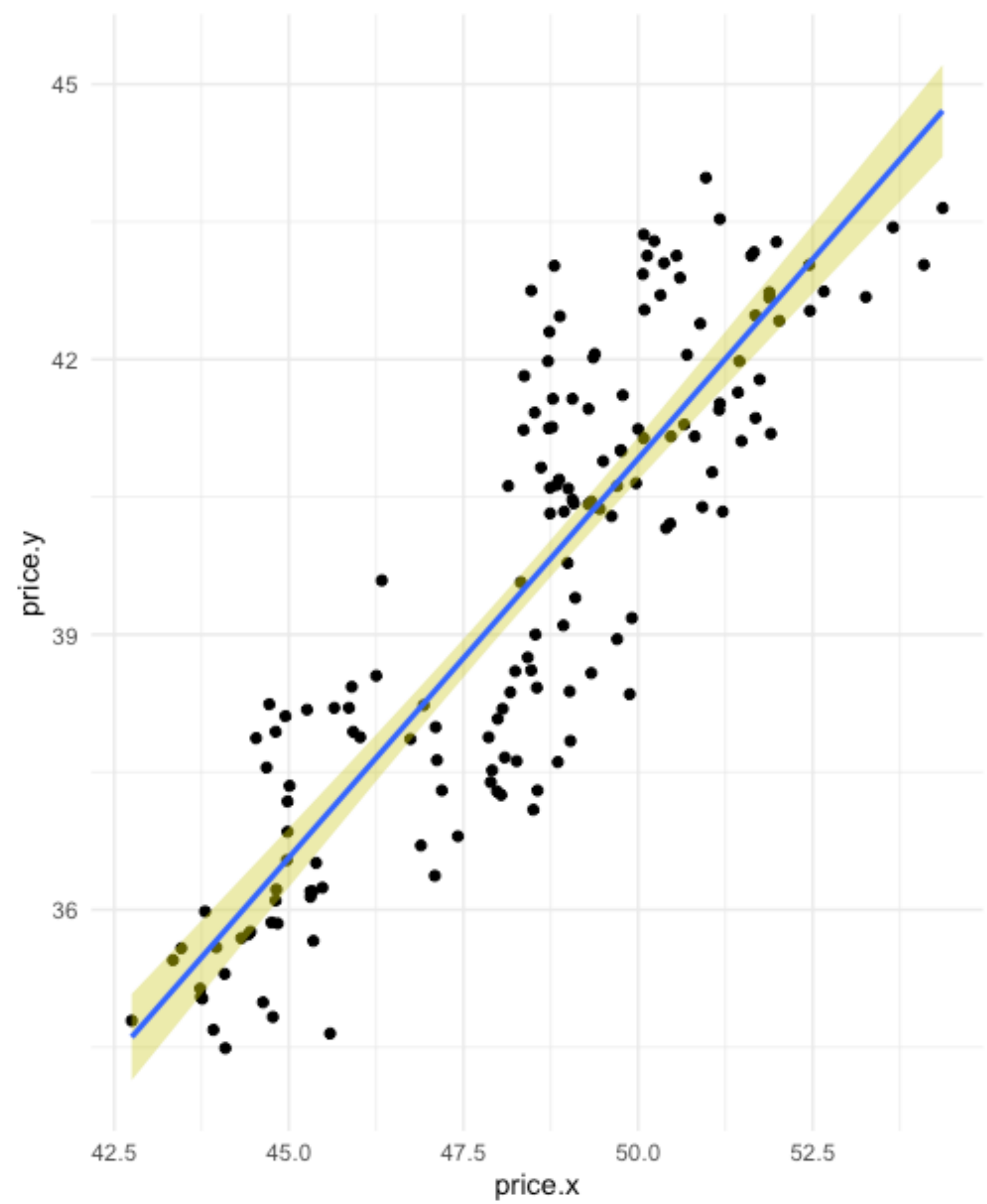
The general form of a linear regression model is:

$$Y_t = \alpha + \beta X_t + \varepsilon_t$$

This model is the starting point for understanding the relationships between time series data. In financial markets, particularly in **pair trading**, we apply this regression to pairs of stocks to determine whether they share a stable, long-term relationship. However, when both time series are **non-stationary** (i.e., they exhibit trends over time), we need to employ **cointegration** analysis to assess whether a common equilibrium exists between them.

The **residuals** from the regression are critical. If they are **stationary**, it indicates that the series will revert to the equilibrium over time, making this the basis for trade signals. Residuals are also used to calculate the **z-score**, a key metric for identifying significant divergences in price and predicting potential reversions to the mean.

By utilizing this base statistical model, we can systematically quantify relationships, test for cointegration, and make informed predictions about future price movements. This model underpins not only financial trading strategies but also a wide array of economic forecasting applications.



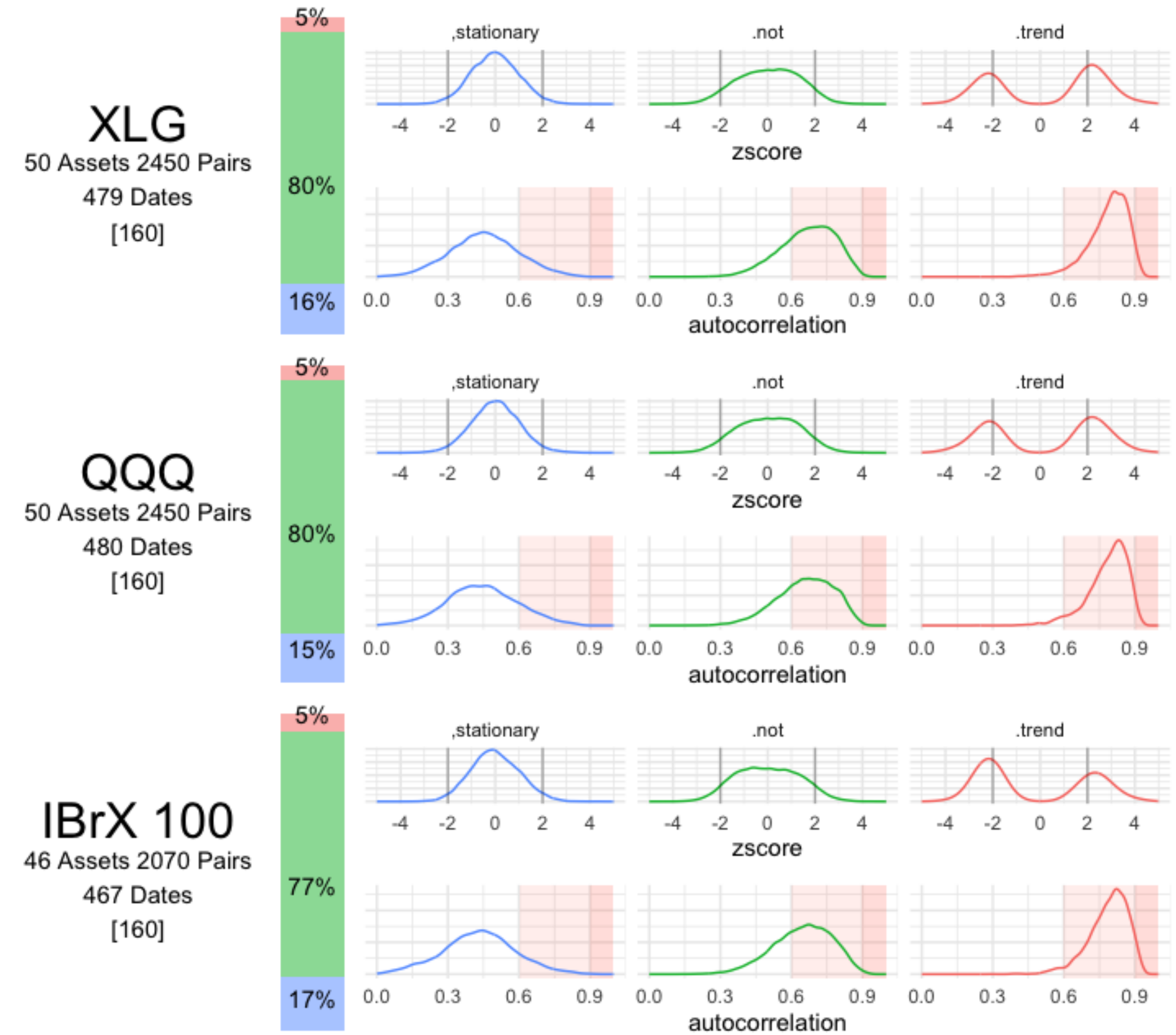
cointegration tests applied to several indexes

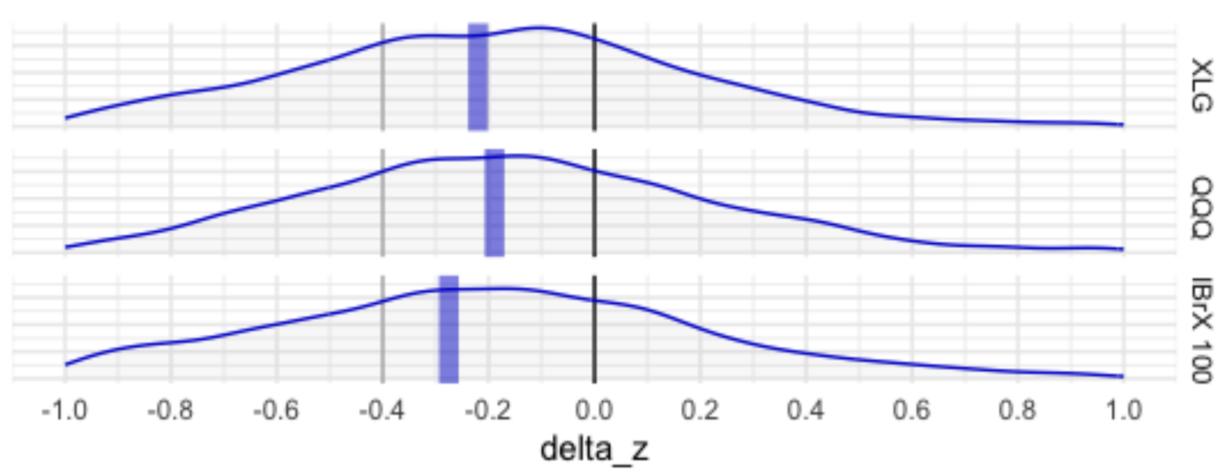
In this project, stationary tests were conducted on pairs of stocks derived from three different financial indexes XLG, QQQ and IBXX to assess the viability of a cointegration-based trading strategy. The primary objective was to evaluate key output variables such as the **p-value**, **z-score**, **auto-correlation** and **delta-z** (the difference between opening and closing prices). By examining these metrics, we aimed to uncover the underlying similarities and slight differences between different indexes. The p-value provided insight into the statistical significance of cointegration group. Auto-correlation was used to confirm patterns and persistence in price movements and delta z offered a comparative measure of daily price shifts. Together, these analyses allowed for a nuanced understanding of how different index pairs behave and their suitability for pair trading strategies.

The **concept of cointegration** in time series analysis was first developed by Robert F. Engle and Clive W.J. Granger. In the early 1980s, Engle and Granger independently introduced the notion of cointegration, which addresses the issue of spurious correlation in non-stationary time series data.

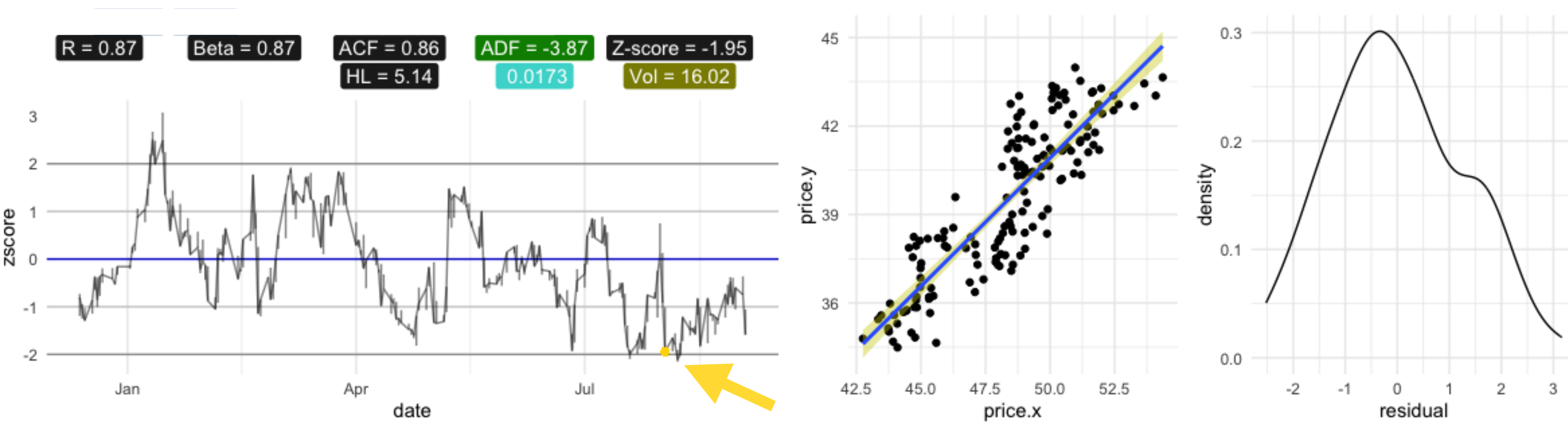
Robert Engle is an economist and Nobel laureate, known for his work on time series analysis and econometrics. Clive Granger, also an economist and Nobel laureate, made significant contributions to the field of econometrics, including his work on cointegration.

Engle and Granger's research on cointegration has had a profound impact on **econometrics**, financial modeling, and economic analysis. Cointegration is widely used to study the long-term relationships and equilibrium behavior among non-stationary time series variables, allowing for more accurate modeling and forecasting of economic and financial data.





execution plan



In econometrics, cointegration is a powerful statistical technique used to identify long-term relationships between two or more time series that are individually non-stationary but exhibit a stable relationship over time. In the context of financial markets, cointegration is particularly valuable for pair trading, where the prices of two assets may diverge in the short term but maintain a consistent equilibrium over the long term. This approach allows us to exploit temporary deviations from this equilibrium, generating profitable trading signals.

To identify cointegrated pairs, we first apply a linear regression model between two time series, say stock A and stock B.

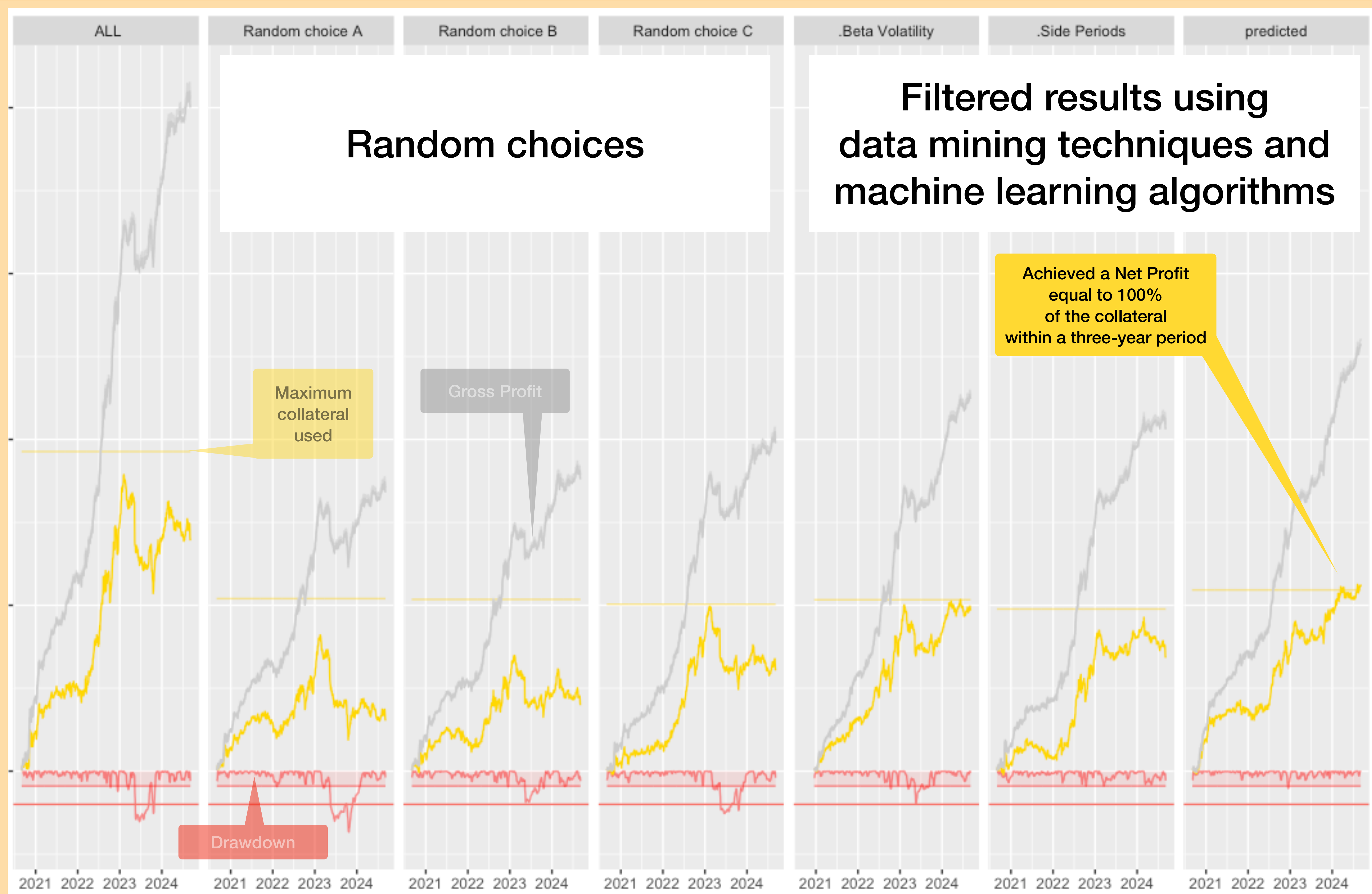
If the residuals are stationary, it indicates that the two time series are cointegrated. This implies that while stock prices may diverge in the short run, they will eventually revert to a common equilibrium.

To verify the presence of cointegration, we apply statistical tests such as the Augmented Dickey-Fuller (ADF) test to the residuals. The ADF test determines whether the residuals are stationary, indicating cointegration. If the test confirms stationarity, we then use the residuals to calculate the z-score, which provides a measure of the deviation from the mean of the residuals.

A z-score exceeding 2 standard deviations is interpreted as a signal to enter a trade, as it represents a significant divergence from the equilibrium, with an expectation that the prices will eventually revert to the mean. When the z-score returns to the mean, the position is closed, capturing the profit from this reversion.

This execution plan, grounded in cointegration and linear regression, forms the basis for generating profitable signals in the pair trading strategy.

backtests results



contact

Marcos Lima

Data Scientist | Quantitative Analyst Researcher

l1m4o981@gmail.com

linkedin.com/in/mkslima

github.com/l1m40/QuantMetricMarketInsights

About Me

With a solid foundation in data analytics and a deep understanding of financial markets, I have developed this automated trading system using cointegration and econometrics methods to optimize pair trading strategies. My expertise in R programming allows me to manage, transform, and model complex datasets, while leveraging advanced machine learning algorithms to enhance predictive accuracy. I also specialize in building interactive data visualizations using Shiny, providing real-time insights and actionable analytics.

This project reflects my strong analytical skills and ability to apply quantitative methods to financial markets. Feel free to reach out for more information or collaboration opportunities.