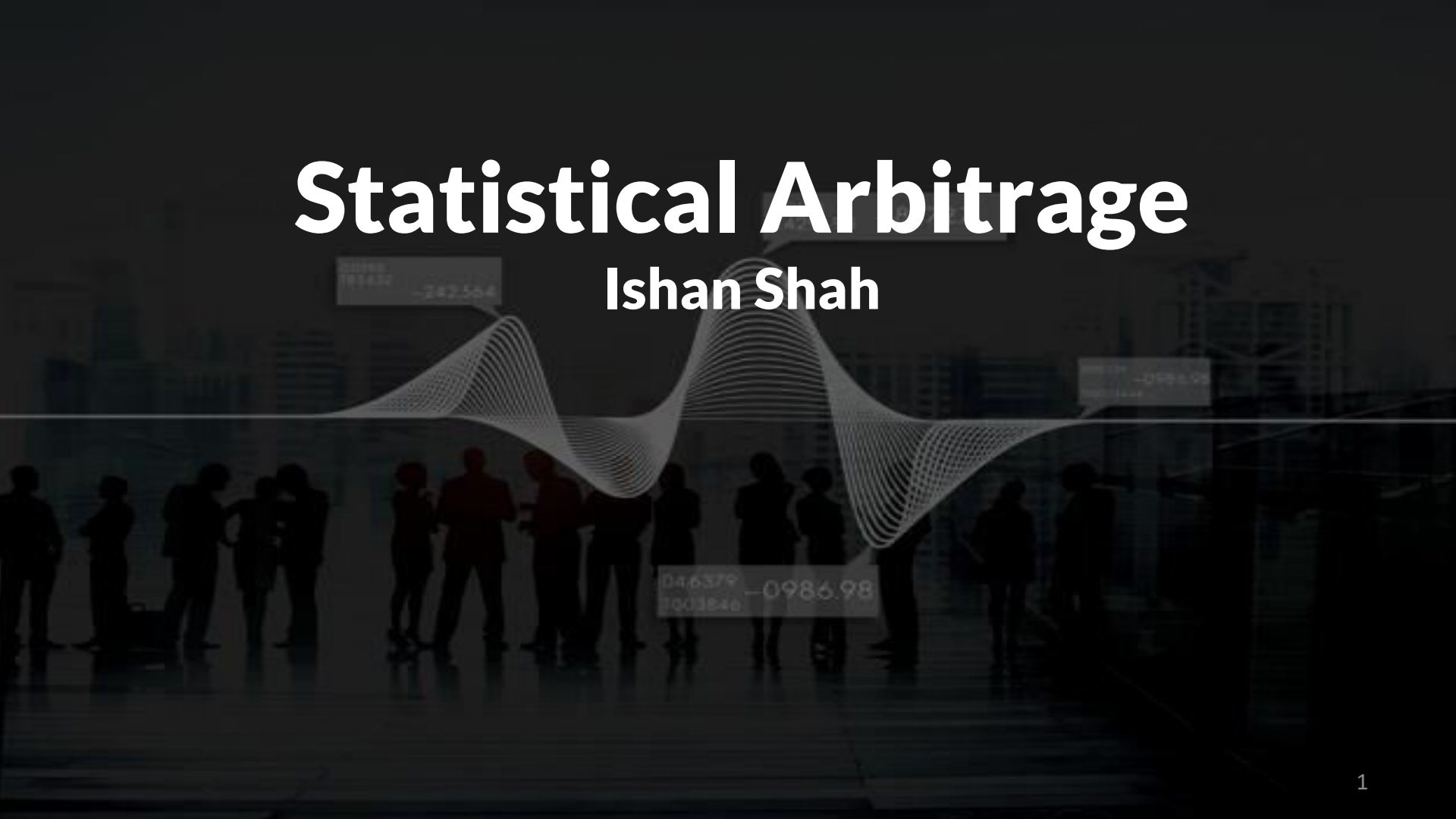
Statistical Arbitrage

Ishan Shah



Outline

Part I

- Stationarity
- Math behind ADF Test
- ADF Test
- Mean-reversal Strategy

Part II

- Cointegration
- Cointegration Test
- Pairs Trading Strategy
- Cointegration vs. Correlation
- Pairs Selection
- Risk Management

Types of Statistical Arbitrage Strategies

- **Directional trading**

  Dependent on single instrument.

  Example: Corn Futures, Gold Futures, etc.

- **Pairs trading, triplets and other cointegrated trading**

  Relative value of 2, 3 or more instruments.

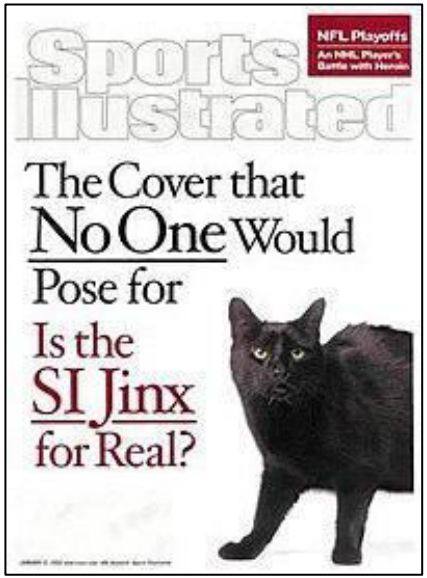
  Example: Google vs. Facebook

Stationarity

Stationarity Analogy

Daniel Kahneman cited a famous example:

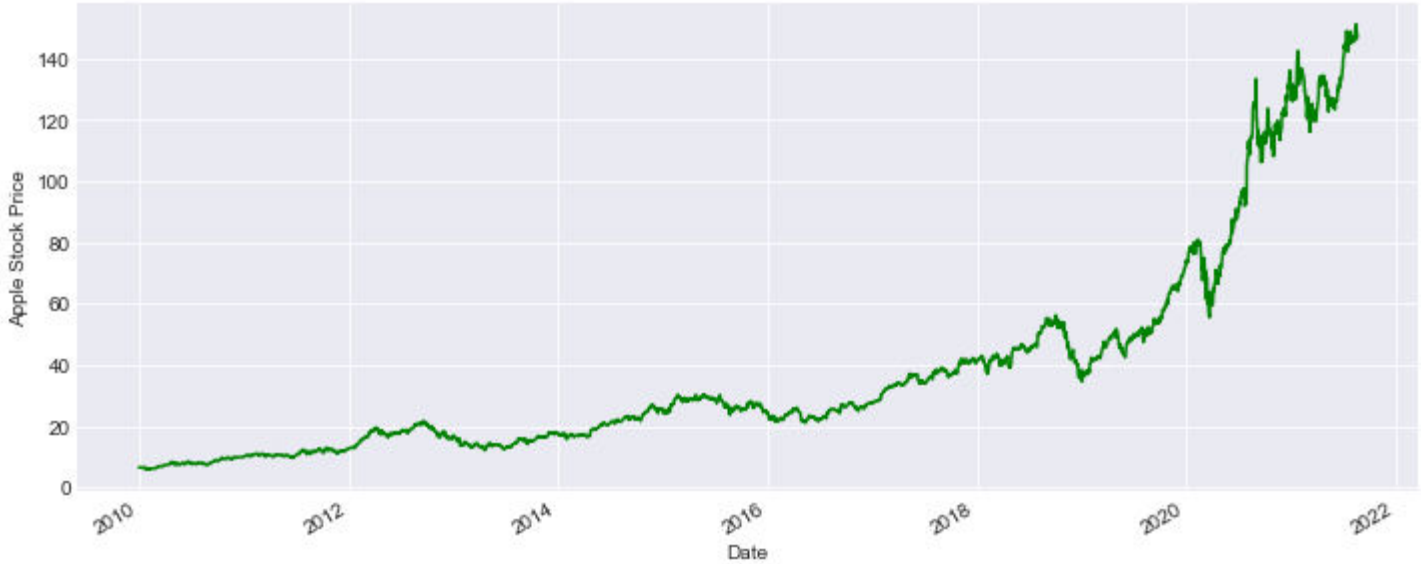
The “Sports Illustrated jinx,” which is the claim that *“an athlete whose picture appears on the cover of the magazine is highly probable to do badly in the upcoming season”*.



The magazine cover for the January 21, 2002 issue of *Sports Illustrated*

Is this stationary?

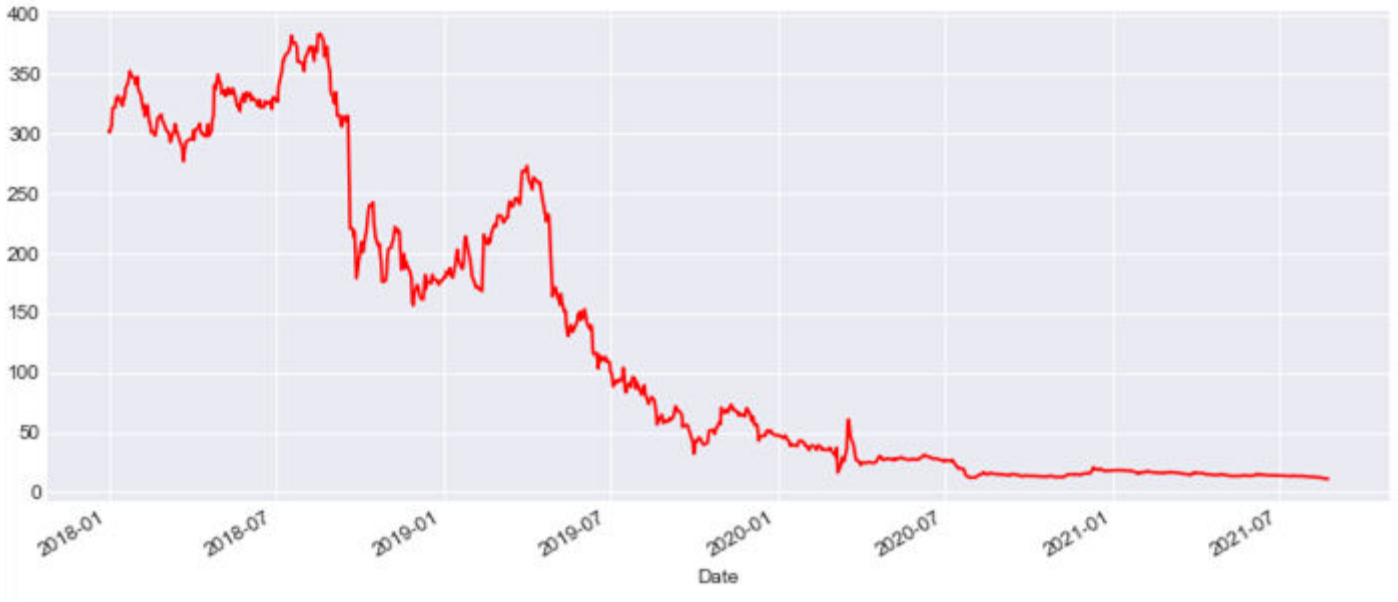
Apple Inc.



Source: Yahoo Finance

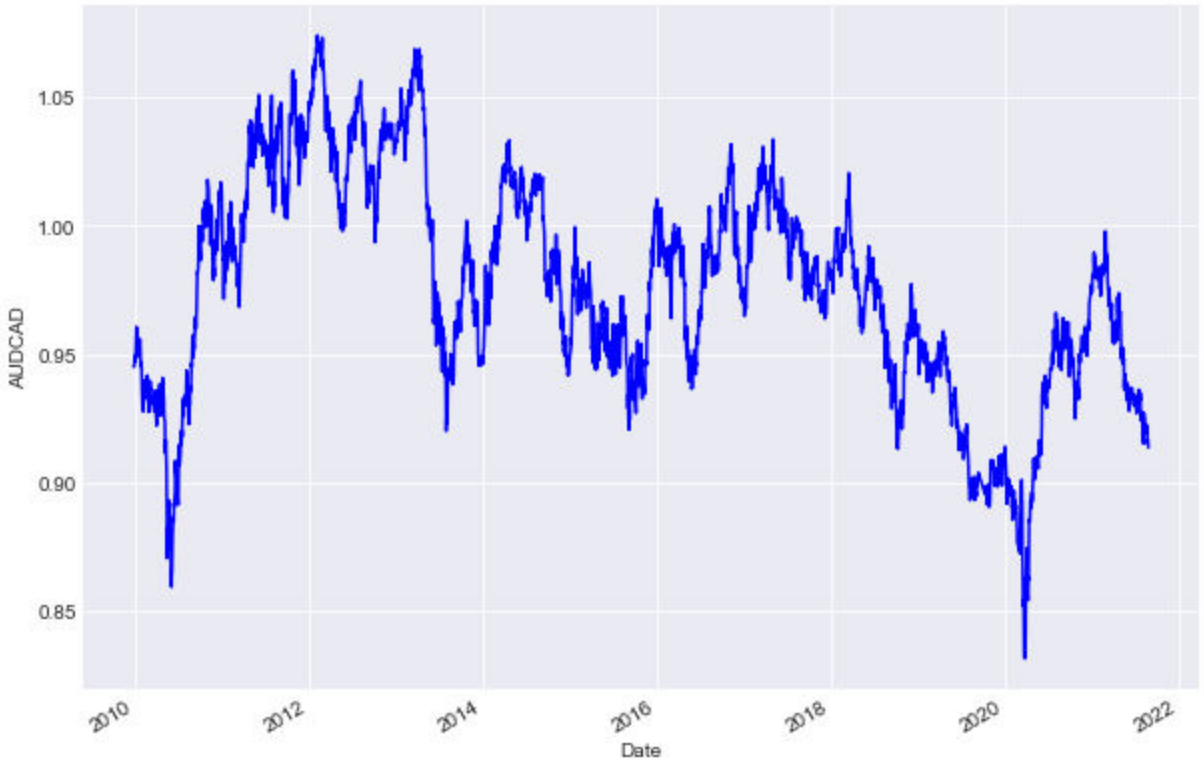
Is this stationary?

Yes Bank



Source: Yahoo Finance

Is this stationary?



Source: Yahoo Finance

What is Stationarity?

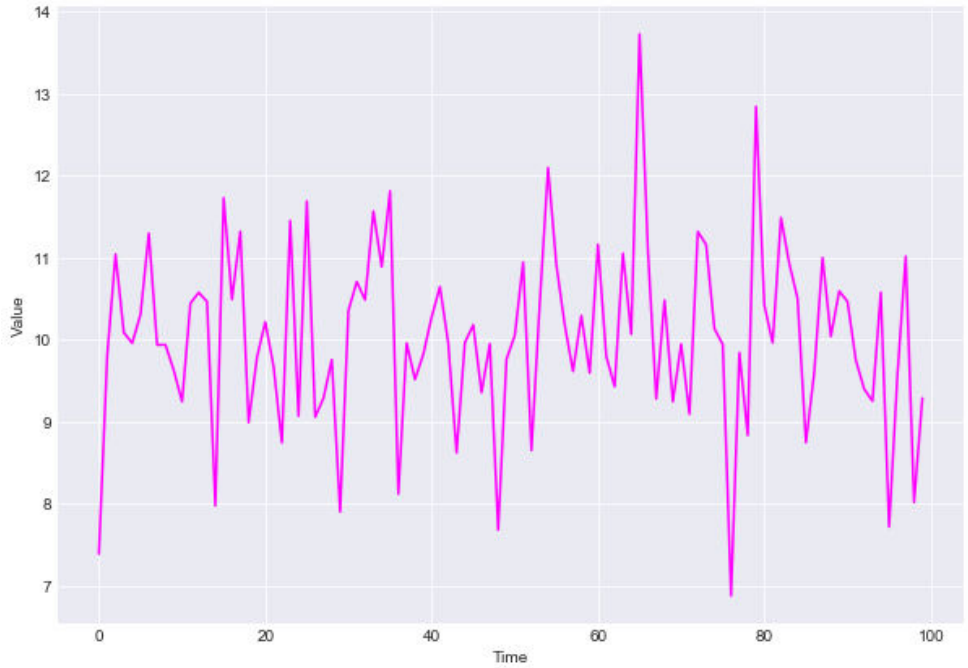
A price series is stationary if it doesn't deviate much but stays around the mean.

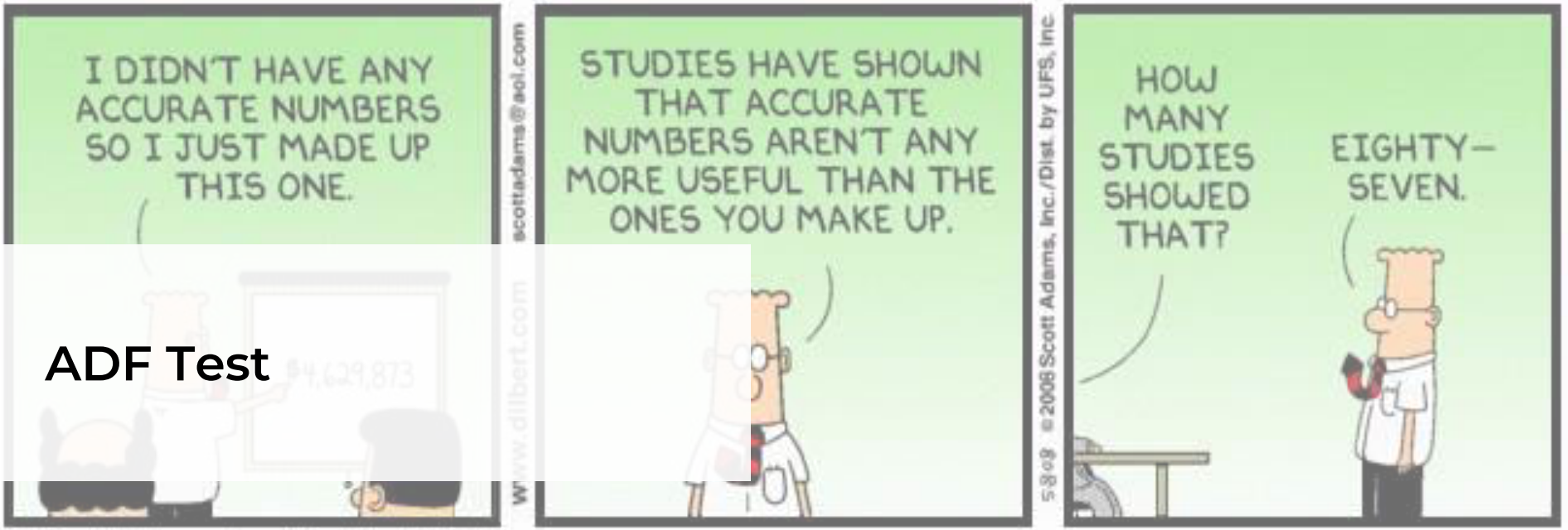
Stationarity

- Random walk is not stationary
- It only applies to long term properties of a price series
- Life is not easy!
- It is difficult to find naturally occurring price series which are stationary

Random walk is not stationary:
https://www.mit.edu/~kardar/teaching/projects/chemotaxis(AndreaSchmidt)/random.htm

Which Strategy will Work on Stationary Time Series?




I DIDN'T HAVE ANY ACCURATE NUMBERS SO I JUST MADE UP THIS ONE.
STUDIES HAVE SHOWN THAT ACCURATE NUMBERS AREN'T ANY MORE USEFUL THAN THE ONES YOU MAKE UP.
HOW MANY STUDIES SHOWED THAT?
EIGHTY-SEVEN.
scottadams@aol.com
www.dilbert.com
5809 ©2008 Scott Adams, Inc./Dist. by UFS, Inc.
© Scott Adams, Inc./Dist. by UFS, Inc.

ADF Test

Testing for Stationarity

- **That's an intuitive definition, but how to check for this statistically?**
- **There is a convenient statistical test:**
  - **Augmented Dickey-Fuller (ADF) Test**

Can’t we just run a backtest on the trading strategy directly and be done with it? Why do we need a statistical test?

Stationarity Test: ADF (1 of 2)

- **We can describe the current price changes from the past price data using a linear model**

  $\Delta p(t) = \lambda p(t-1) + \mu + \beta_t + \alpha_1 \Delta p(t-1) + \ldots + \alpha_k \Delta p(t-k) + \epsilon t$

  p= price of the instrument

  $\Delta p(t) = p(t) - p(t-1)$,

  $\Delta p(t-1) = p(t-1) - p(t-2)$, *and so on.*

  λ *= regression coefficient*

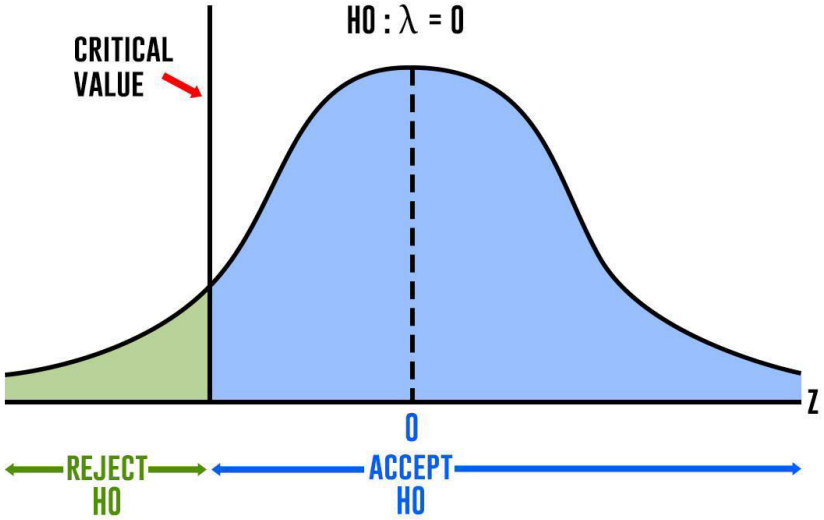
- **H_0: $\lambda = 0$** **Not Stationary**
- **H_a: $\lambda < 0$** **Stationary**
- **Intuition:**

  If the hypothesis $\lambda = 0$ is rejected -

  Next move ($\Delta p(t)$) --> Current level ($p(t-1)$)

  Hence, it is stationary.

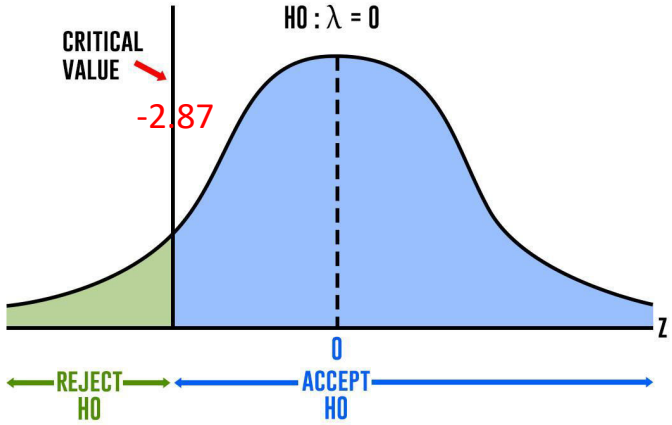
Stationarity Test: ADF (2 of 2)



Probability	Critical Value
90%	-2.59
95%	-2.87
99%	-3.44

Stationarity Test: Quiz

1. **Is t-stat value of -3.05 stationary with 95% confidence level (critical value: -2.87)?**
   - Yes/No
2. **Is t-stat value of -2.40 stationary with 95% confidence level (critical value: -2.87)?**
   - Yes/No



Testing for Stationarity in Python (1 of 2)

- **Read in AUDCAD.csv file**

  data = pd.read_csv('file_name.csv', index_col=0)

- **Plot the time series**

  data.Close.plot(figsize=(8,4))

- **Check for Stationarity using adfuller method available in statsmodels.tsa.stattools package**

  help(adfuller)

Testing for Stationarity in Python (2 of 2)

- **Set *maxlag = 1* (Assuming short-range serial correlation)**

  Perform ADF Test

  adf = adfuller(data.Close, maxlag = 1)

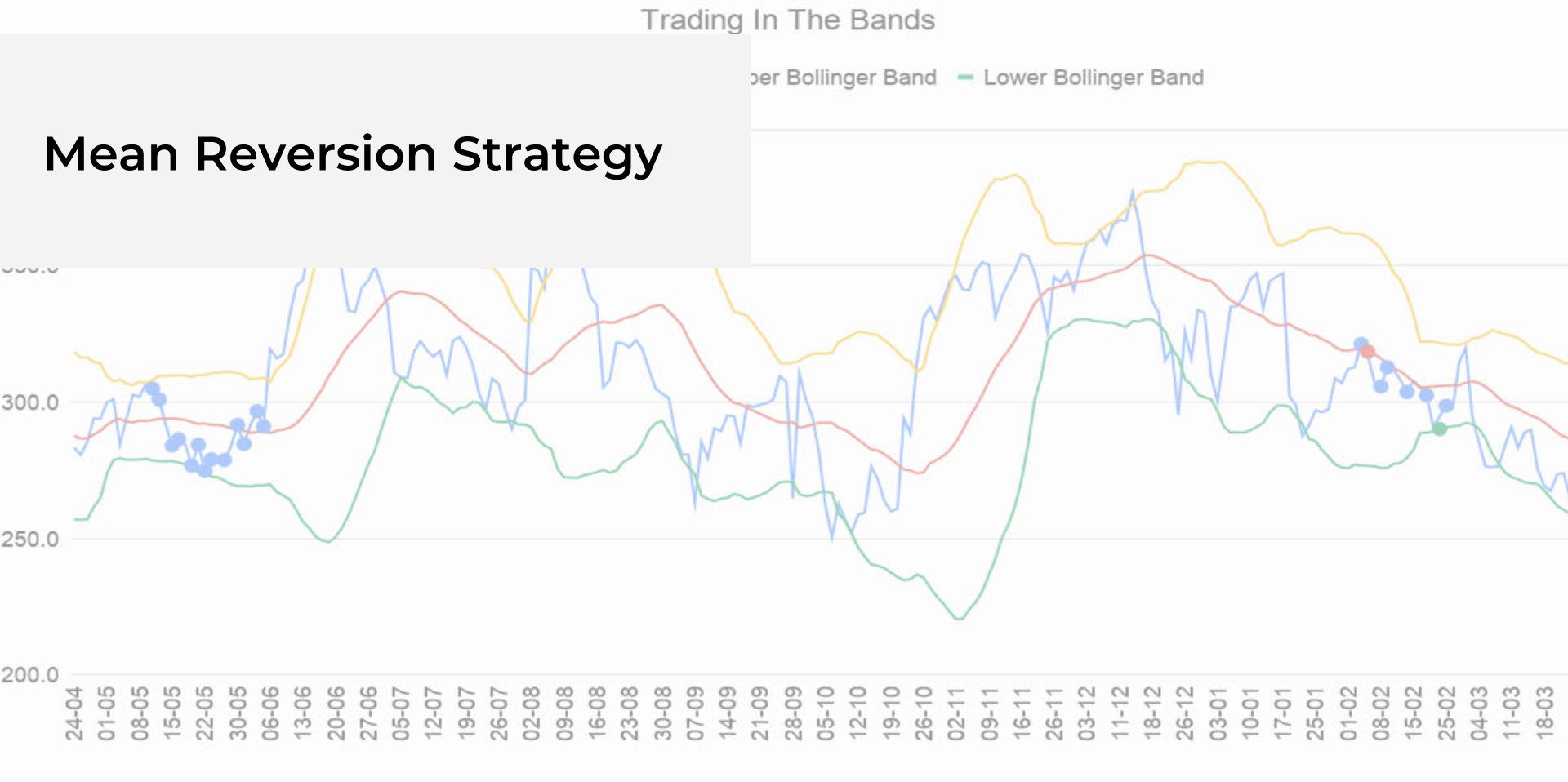
  (Advanced topic: To find optimal maxlag, use AIC/BIC.)

- **Test statistic: *adf[0]***

- **<u>Conclusion:</u>**

  Since t-stat < -2.59, AUDCAD is stationary with more than 90% certainty

Mean Reversion Strategy



Mean Reversion Strategy in Python

- Read data from CSV file
- Moving average and moving standard deviation
- Upper band and lower band
- Long positions
- Short positions
- Strategy PnL

Mean Reversion Strategy in Python

- **Read data from CSV file**
- **Moving average and moving standard deviation**

  df['moving_average'] = df.Close.rolling(5).mean()

  df['moving_std_dev'] = df.Close.rolling(5).std()
- **Upper band and lower band**

  *df['upper_band'] = df.moving_average + 0.5*df.moving_std_dev*

  *df['lower_band'] = df.moving_average – 0.5*df.moving_std_dev*

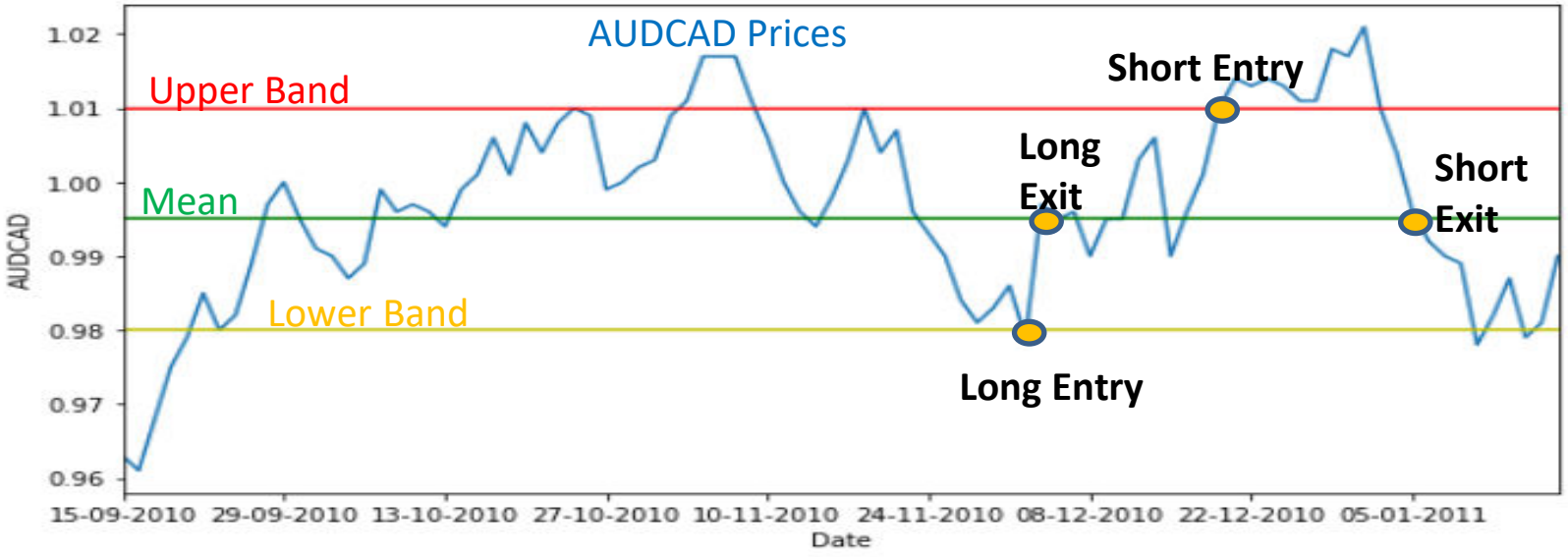
How to select lookback period (5) and width of Bollinger bands (0.5)?

1. Optimize in training set
2. Set lookback to some multiple of half-life

https://github.com/QuantInsti/EPAT/blob/master/Statistical%20Arbitrage/optimization.py

https://github.com/QuantInsti/EPAT/blob/master/Statistical%20Arbitrage/half_life.py

Bollinger Bands: AUDCAD



Note: The straight line for mean, upper band and lower band is for illustration purpose only. In the strategy, rolling mean, rolling upper band and rolling lower band is used.

Data source: FXCM

Mean Reversion Strategy in Python

Day	df.Close	df.moving_average	df.lower_band	df.long_entry	df.long_exit
1	98.5	99.5	98	FALSE	FALSE
2	97.5	99.5	98	TRUE	FALSE
3	98.2	99.5	98	FALSE	FALSE
4	99.2	99.5	98	FALSE	FALSE
5	99.6	99.5	98	FALSE	TRUE
6	100	99.5	98	FALSE	TRUE
7	99	99.5	98	FALSE	FALSE

- **Long Entry**

  df['long_entry'] = df.Close < df.lower_band

- **Long Exit**

  df['long_exit'] = df.Close >= df.moving_average

Mean Reversion Strategy in Python

df.long_entry	df.long_exit	positions_long
FALSE	FALSE	NaN
TRUE	FALSE	1
FALSE	FALSE	NaN
FALSE	FALSE	NaN
FALSE	TRUE	**0**
FALSE	TRUE	**0**
FALSE	FALSE	NaN

- **Long Positions**

df['positions_long'] = np.nan

df.loc[df.long_entry,'positions_long'] = 1

df.loc[df.long_exit,'positions_long'] = 0

Mean Reversion Strategy in Python

df.long_entry	df.long_exit	positions_long	Forward Filled (positions_long)
FALSE	FALSE	NaN	NaN
TRUE	FALSE	1	1
FALSE	FALSE	NaN	1
FALSE	FALSE	NaN	1
FALSE	TRUE	**0**	**0**
FALSE	TRUE	**0**	**0**
FALSE	FALSE	NaN	**0**

- **Forward Fill NaN Values**

  Carry forward an existing position, whenever the next bar's position has not been predetermined to be 0 or 1

  df.positions_long = df.positions_long.fillna(method='ffill')

Mean Reversion Strategy in Python

- **Short Positions**

  df['short_entry'] = df.Close > df.upper_band

  df['short_exit'] = df.Close <= df.moving_average

  df['positions_short'] = np.nan

  df.loc[df.short_entry,'positions_short'] = -1

  df.loc[df.short_exit,'positions_short'] = 0

- **Carry forward an existing position, whenever the next bar's position has not been predetermined to be 0 or -1**

  df.positions_short = df.positions_short.fillna(method='ffill')

Mean Reversion Strategy in Python

- **Final Positions**

  df['positions'] = df.positions_long + df.positions_short

- **Pnl**

  df['prices_difference'] = df.Close - df.Close.shift(1)

  *df['pnl'] = df.positions.shift(1) * df.prices_difference*

  df['cumpnl'] = df.pnl.cumsum()

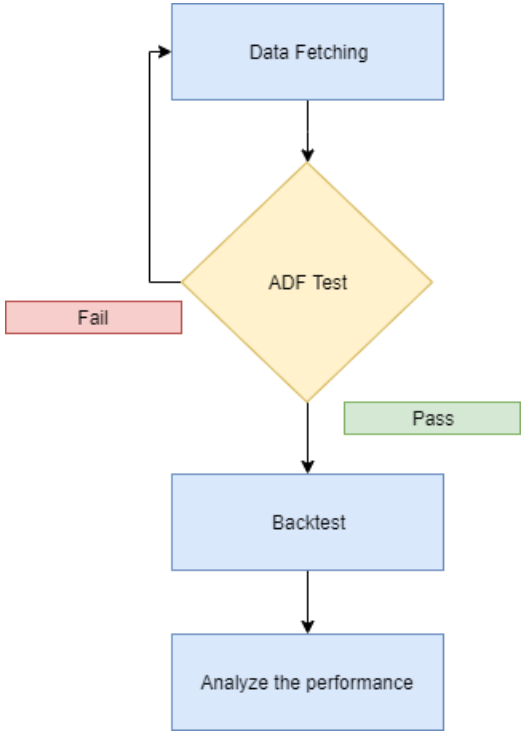
- **Returns**

  df['percentage_change'] = df.Close.pct_change()

  *df['strategy_returns'] = df.positions.shift(1) * df.percentage_change*

  df['cumulative_returns'] = (df.strategy_returns+1).cumprod()

Summary



Pairs Trading

Historical Background: Pairs Trading

- First developed and used in the mid 1980s by Nunzio Tartaglia's quantitative group at Morgan Stanley.
- Pair Trading is a "*contrarian strategy*" designed to harness mean-reverting behavior of cointegrated instruments.
    - Question: Why Contrarian?
- David Shaw, founder of D.E. Shaw & Co., left Morgan Stanley and started his own "Quant" trading firm in the late 1980's dealing mainly in pair trading.

Cointegration

- A portfolio of two or more instruments such that the portfolio is stationary. Then, the instruments in the portfolio are said to be cointegrated
- Mean Reversion Strategy
  - *Buy the portfolio when its value is low*
  - *Sell the portfolio when its value is high*

Pairs Trading

- EWA: iShares MSCI Australia ETF

  The iShares MSCI Australia ETF seeks to track the investment results of an index composed of Australian equities.

- EWC: iShares MSCI Canada ETF

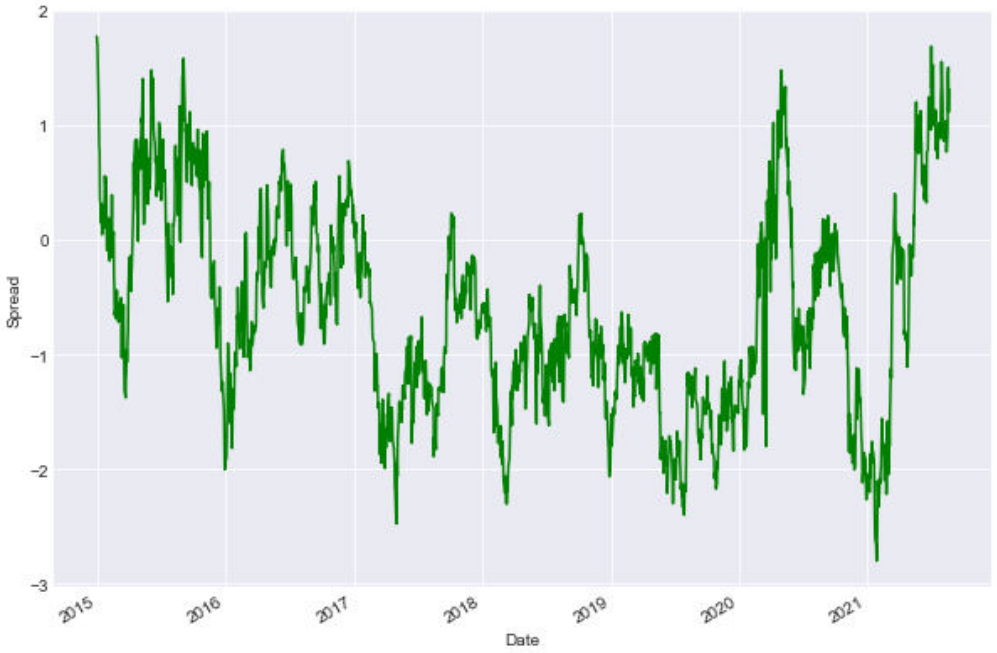
  The iShares MSCI Canada ETF seeks to track the investment results of an index composed of Canadian equities

Source: https://www.ishares.com/us/products/239607/ishares-msci-australia-etf

Example of Prices of EWA and EWC

Data source: Yahoo finance

Example of Cointegrated Price Series



Spread = EWC – 1.45 * EWA

Strategy – buy low and sell high!

Data source: Yahoo finance

Cointegration

- Any arbitrary combination of the instruments are not cointegrated
- Only some combinations of the instruments are cointegrated
- Linear regression to find the hedge ratio

```
import statsmodels.api as sm

model = sm.OLS(df.EWC.iloc[:90], df.EWA.iloc[:90])

model = model.fit()

print(model.params[0])
```

How to select the lookback period (90 days)?

Cointegration Test in Python

- **Consider two ETFs as X (EWA) and Y (EWC)**
  - **Hedge Ratio: Linear Regression *X and Y***
  - **Spread: *Y – Hedge Ratio*X***
  - **Stationarity Test (ADF test) on the spread**
- **Test statistic: *adf[0]***
- **<u>Conclusion:</u>**

  Since t-stat < -2.56, EWA and EWC is cointegrated with more than 90% certainty

Mean Reversion Strategy

- **Compute**
  - **Spread: Y – Hedge Ratio * X**
  - **Mean of the Spread**
  - **Standard Deviation of the Spread**
  - **Upper Band and Lower Band**

Mean Reversion Strategy

- **Entry conditions**

  Long Entry:

  If spread < lower band then

  buy 1 share of Y

  sell hedge ratio share of X

  Short Entry:

  If spread > upper band then

  sell 1 share of Y

  buy hedge ratio share of X

- **Exit conditions**
- **PnL**

Cointegration Vs. Correlation

Cointegration vs. Correlation

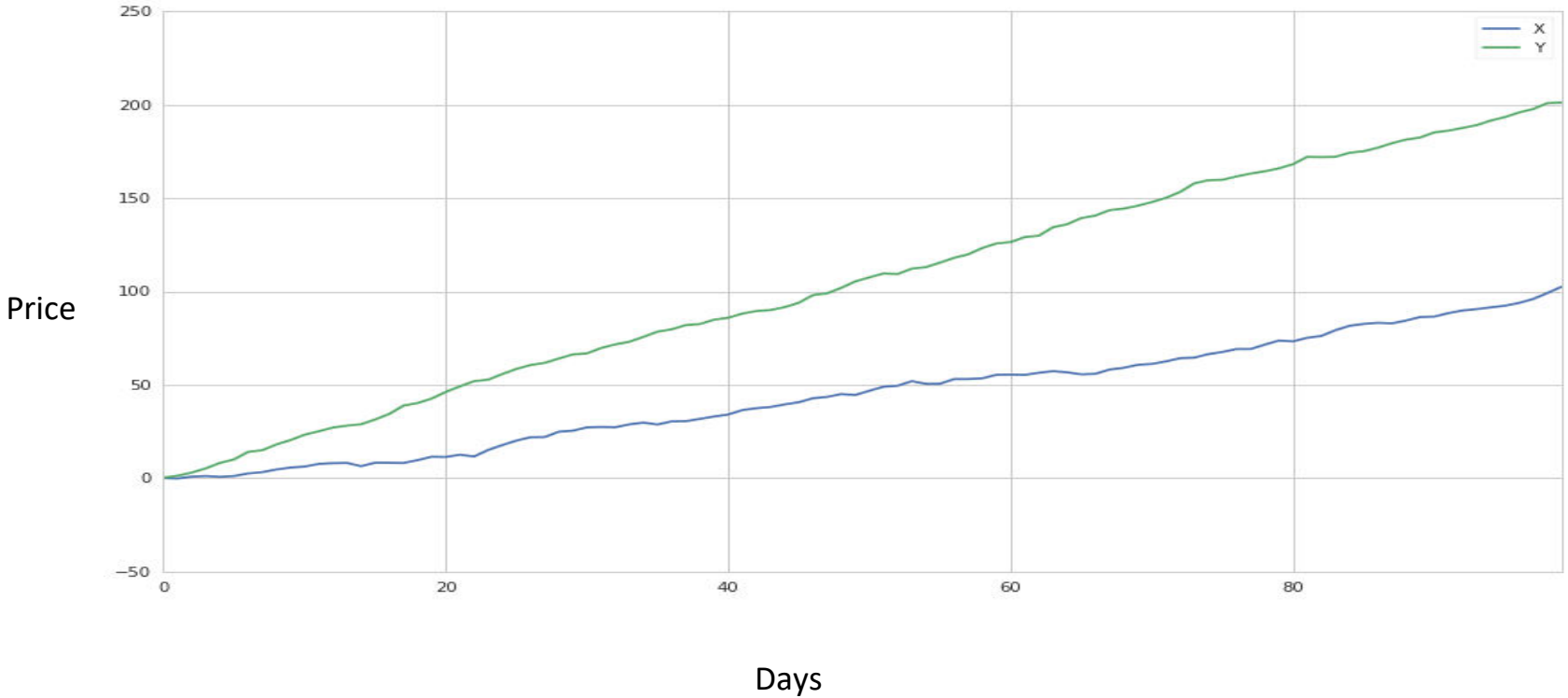
Cointegration

- It is whether the spread between two instruments is stationary
- Long-term

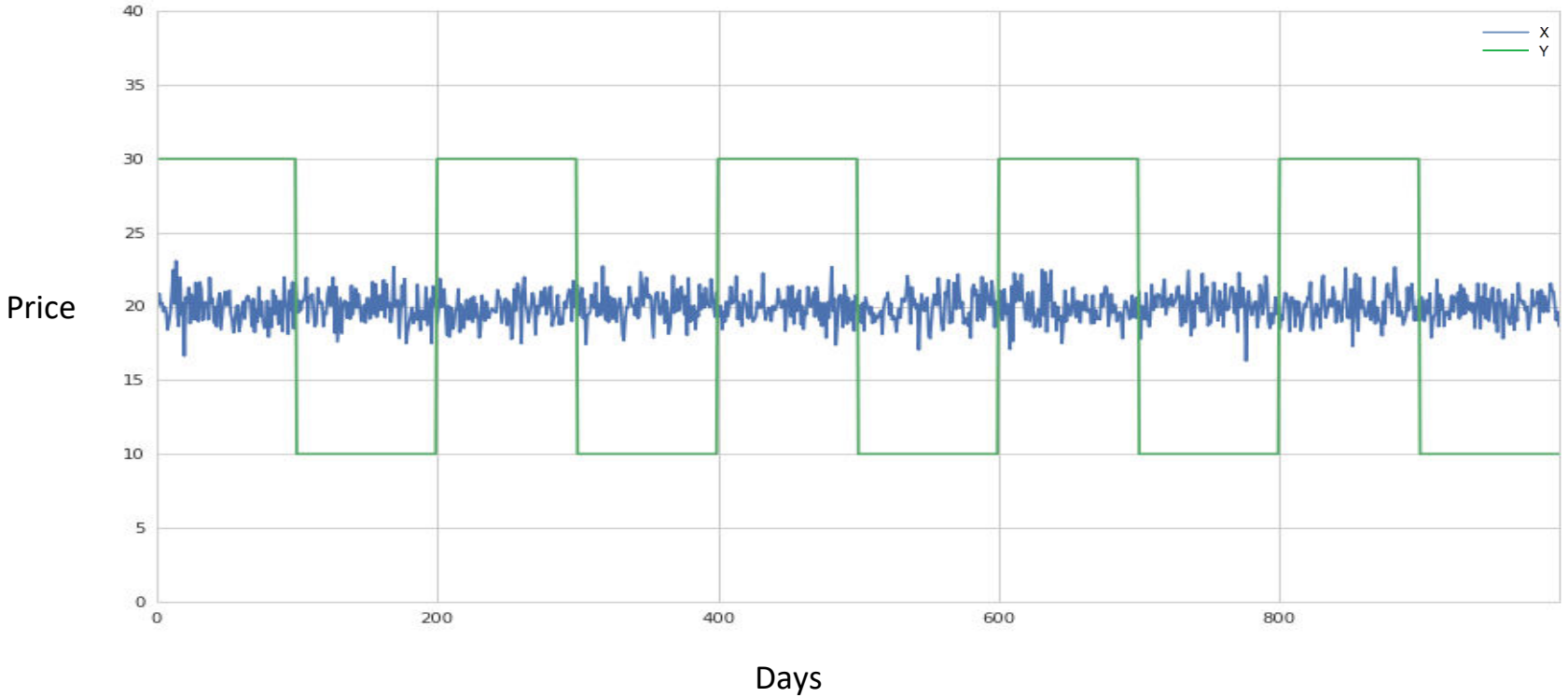
Correlation

- It is whether the two instruments returns move in same direction
- Short-term or long-term

Correlation without Cointegration



Cointegrated without Correlation



Pairs Selection

Why Pair Trading Works?

- Cointegrated pair are assumed to be mean reverting in nature
- Once the spread diverges from its mean, the probability of reversal increases
- With the chosen pair from same industry and sector, extreme divergence from mean is rare

The Concept (1 of 2)

- Pair Selection Criteria
  - Qualitative Selection (ETFs & Currencies)
    - Exposed to common economic factors
  - Qualitative Selection (Stocks)
    - Same Sector
    - Similar Market Capitalization
    - Similar Ratios

    Note: Many choices of pairs but easy to lose cointegration
- Cointegration Check
    - ADF Test

The Concept (2 of 2)

- Example 1: Both (EWA) Australia and (EWC) Canada are dominated by commodities stocks
- Example 2: ACC and Ambuja

Criteria	ACC	Ambuja
Sector	Cement & Cement Products	Cement & Cement Products
Market Capitalization (In INR Cr.)	32,805	53,235
ROE (3 Years)	9.88%	8.74%

Risk Management

Case Study: LTCM

Risk Management

- **Statistical Arbitrage is not a risk-free strategy**
- **Rather than converging, the spread can begin to diverge (drift apart)**
  - *The spread picks up trend rather than mean-reverting and the cointegration is broken*
- **An event in a security can trigger extreme movement in the spread**

Strict risk management is required to handle adverse situations once the mean-reverting behavior is invalidated

Risk Management Ways

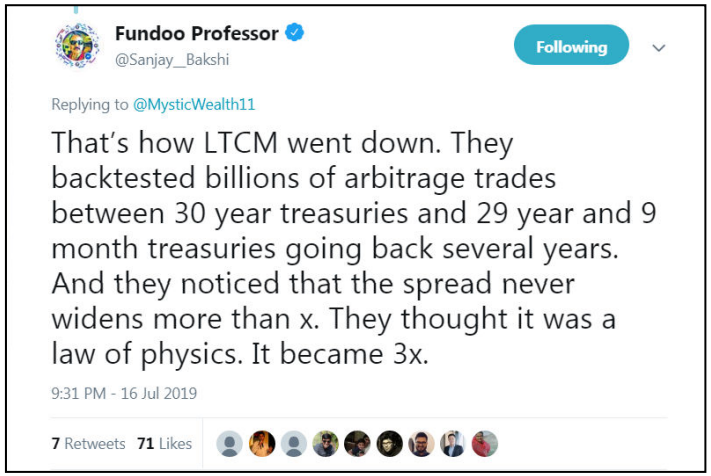
- Define strict stop-loss and profit mechanism before entering trade
- It's a good practice to allocate fund to different pair portfolio rather than one single trade
- Combine with momentum strategies

How to place the stop-loss in these strategies?

Case Study: LTCM

Recipe for disaster

- **Overreliance on backtest**
- **High leverage**



Summary

Summary

- ✔ Stationarity and Cointegration are important for profitable mean reversion strategy
- ✔ ADF Test is used to determine whether two price series is stationary/cointegrated or not
- ✔ Stock pairs are quite unstable w.r.t. cointegration
- ✔ ETFs and currency pair are good candidates for pair trading

Further Readings

- Index arbitrage strategy
- Johansen test
- Common pitfalls in backtesting
- Half-life

Bibliography & References

- Ernest P. Chan “Algorithmic trading winning strategies and their rationale”
- Gatev, Evan, William N. Goetzmann, and K. Geert Rouwenhorst, “Pairs Trading: Performance of a Relative-Value Arbitrage Rule,” *Review of Financial Studies* (2006): 797-827
- Vidyamurthy, Ganapathy, *Pairs Trading: Quantitative Methods and Analysis* (New Jersey: John Wiley & Sons, Inc., 2004)
- Wooldridge, Jefferey M., *Introductory Econometrics, A Modern Approach, Third Edition* (Ohio: Thomson South-Western, 2006)
- Screener.in

Keep in Touch

- Email
- LinkedIn
- Quora

Additional Readings

- **Half-life**
  https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1404905
  https://en.wikipedia.org/wiki/Ornstein%E2%80%93Uhlenbeck_process
- **The Cointegration Alpha**
  https://papers.ssrn.com/sol3/papers.cfm?abstract_id=315619
- **Statistical Arbitrage in the U.S. Equities Market**
  https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1153505
- **Optimal Mean Reversion Trading with Transaction Costs and Stop-Loss Exit**
  https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2222196
- **Implementation of Pairs Trading Strategies**
  https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1594066
- **Quantitative Spread Trading on Crude Oil and Refined Products Markets**
  https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1932471
- **Pairs Trading on International ETFs**
  https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1958546

Additional Readings

- **Pairs Trading: Performance of a Relative Value Arbitrage Rule**
  https://papers.ssrn.com/sol3/papers.cfm?abstract_id=227415
- **Cointegration Trading with Log Prices vs. Prices**
  https://www.jstor.org/stable/4480875?seq=1#page_scan_tab_contents
  http://epchan.blogspot.com/2013/11/cointegration-trading-with-log-prices.html
- **Johansen Test in Python**
  https://www.quantinsti.com/blog/johansen-test-cointegration-building-stationary-portfolio/
- **A Simplified Approach to Understanding the Kalman Filter Technique**
  https://papers.ssrn.com/sol3/papers.cfm?abstract_id=715301
- **Application of Machine Learning Tools in Predictive Modeling of Pairs Trade in Indian Stock Market**
  https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3159868

Thank you!