Scientific Approach to Market Prediction

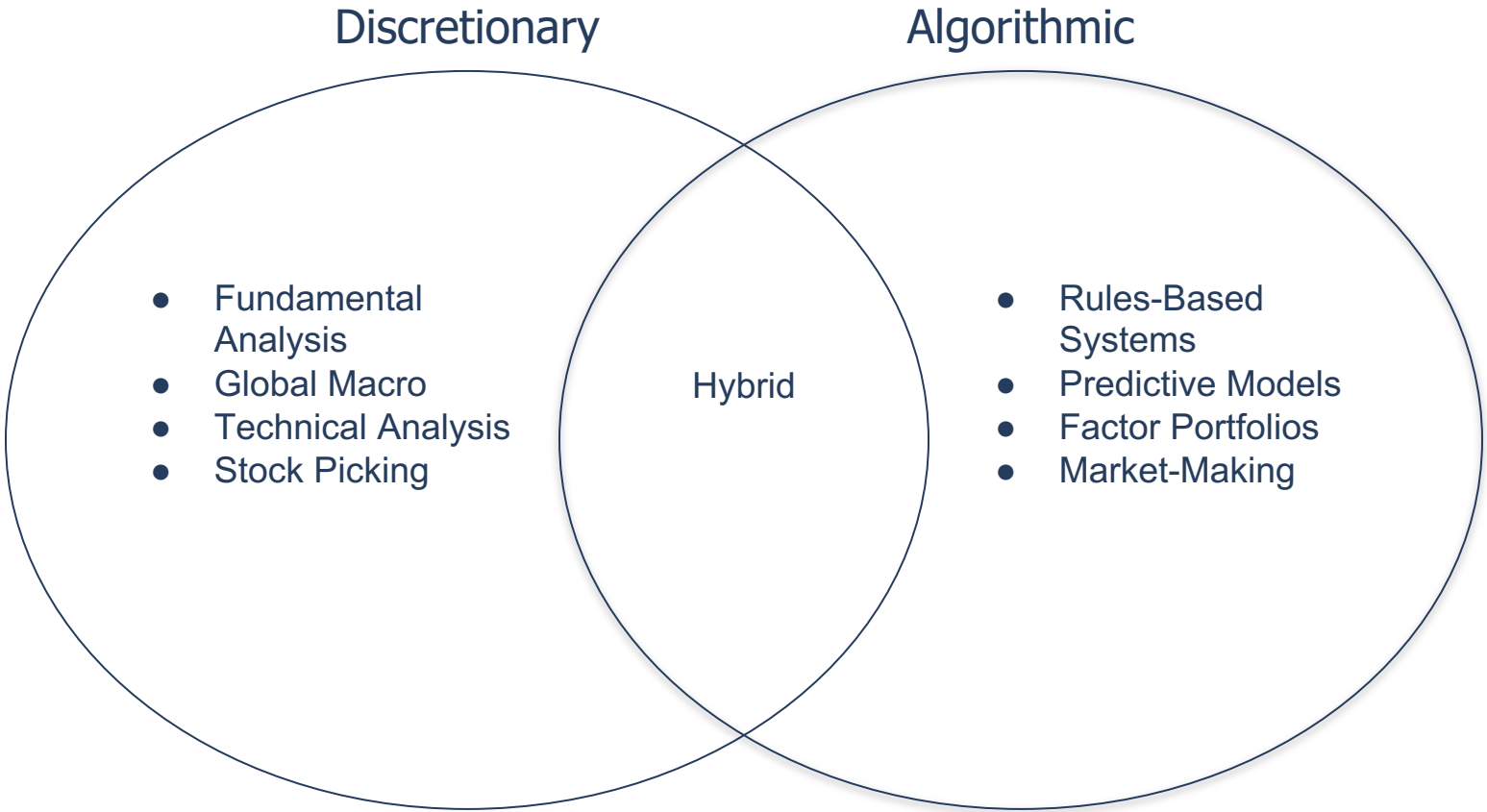
Brian Blandin

About Me

Brian Blandin

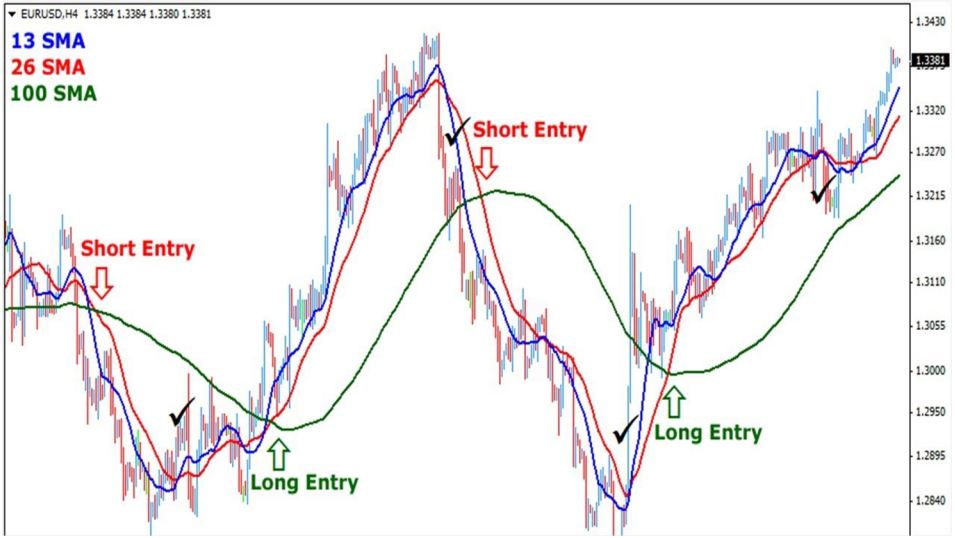
- Co-Founder, Markets Science
- Data Analytics Engineer
- Data Science graduate student
- > 5 years trading experience (equities, options, crypto)

Common Approaches



Rules-Based Systems Overview

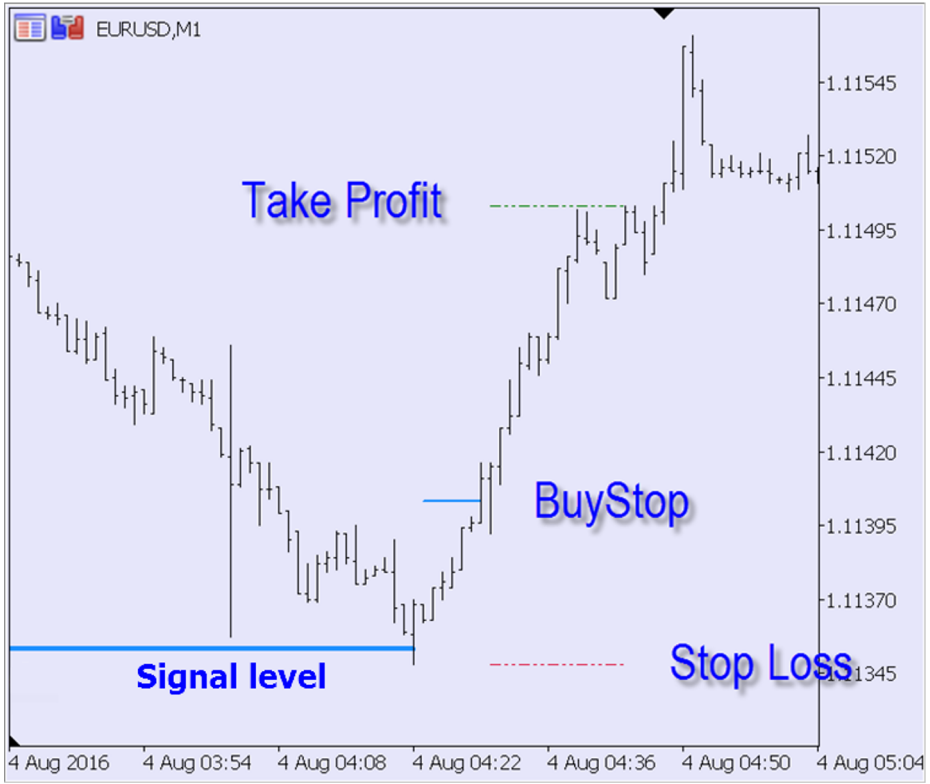
- Logic that takes recent historical data (price or otherwise) and converts to buy/sell signals
- Can typically be written out in pseudo code
- Outputs are often “impulse” signals -- buy or sell
- Can be tweaked into “state”/continuous signals -- position weights



https://www.topdogtradingreview.com/how-to-trade-100-profitable-trading-moving-average-crossover-for.php/

Rules-Based Systems: Pros

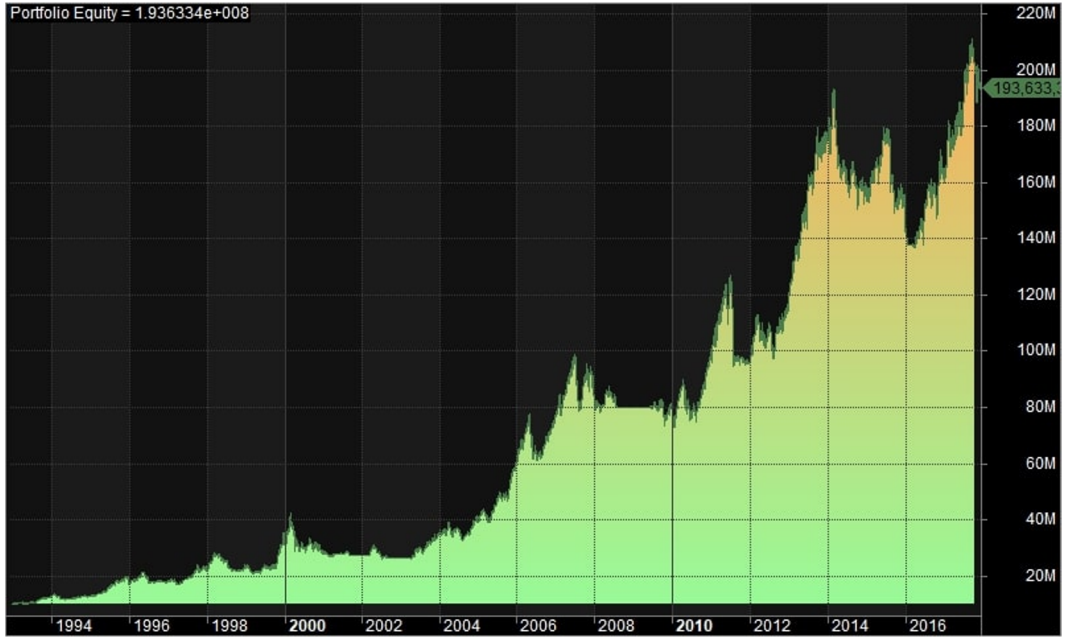
- Easily understandable
  - If the system's conditions are met, a trade's taken
- Available trading software allows creation of most systems
- Easy to troubleshoot
- Lots of books/materials to use as resources



https://www.mql5.com/en/articles/2785

Rules-Based Systems: Cons

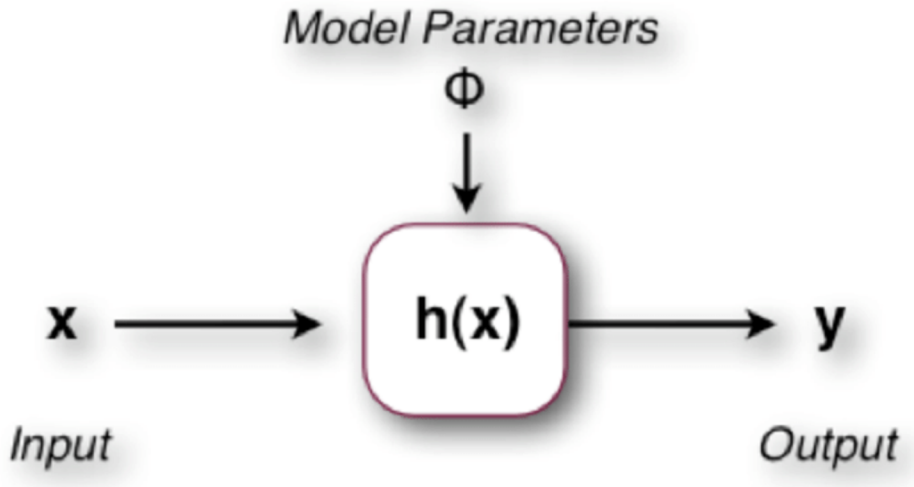
- Overfitting/biases (could be its own presentation!)
- Usually depends on a backtest for evaluation
  - Objective functions commonly based on equity curve
  - Other metrics on trade-by-trade basis
- Typically use hard limits/thresholds
  - Conditions either met or they aren't



https://enlightenedstocktrading.com/backtest-stop-loss-orders-with-amibroker/

Predictive Models Overview

- Attempts to "learn" the relationship between input data and output
- Common approaches:
  - Machine Learning ("AI")
  - Time Series Analysis (GARCH, ARIMA, etc)
- Models are trained on data from the past, assuming that relationships will hold in the future
- Commonly predicted: future prices, future returns, future price direction (up/down)



https://www.researchgate.net/figure/A-machine-learning-algorithm-expressed-as-a-function-h-x-with-input-x-output-y-and_fig1_230691317

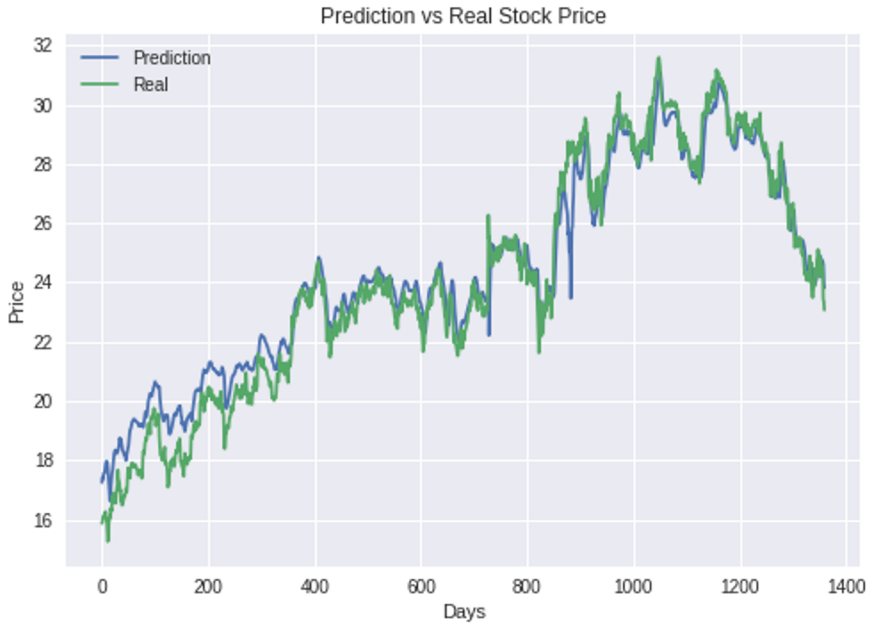
Predictive Models: Pros

- Human trader doesn't need to know/define relationship between inputs+outputs
- Can easily incorporate different input features
- Can detect nonlinear interactions between different data sets
- Objective functions can be applied to predictions or actual trading results
- Most methods provide a non-binary prediction (continuous value or probability)

https://www.datarobot.com/blog/using-machine-learning-to-predict-out-of-sample-performance-of-trading-algorithms/

Predictive Models: Cons

- Overfitting
- Complex models may need more data than available
- Training can be time-consuming / expensive
- “Black boxes”
- Common targets are noisy and difficult to predict



https://towardsdatascience.com/predicting-stock-price-with-lstm-13af86a74944

Commonalities

- Inputs must have predictive value -- garbage in, garbage out
- Data must be representative
  - Adequate history
  - Varied conditions
  - Available at time of trade
- Avoid biases/pitfalls
- Issue: market regimes change
  - Most fitting exercises will benefit the dominant regime at expense of others

A Better Way

- If regime is known, system design is straightforward
  - Trend-following when trending
  - Mean reversion when reverting
  - Volatility selling when volatility contracts (and vice versa)
- Parameters/Algorithm specification matter less when you have the theme broadly correct
- There are typically multiple ways to express a given trade idea

https://www.google.com/url?sa=i&url=http%3A%2F%2Fblog.swipestudio.co%2Fliquidity-buying%2Fhigher-high-and-lower-low-trading-system-donchian-channel-explanation%2F&psig=AOvVaw12swMkw3F3PqcMIfHqfqxM&ust=1626444247643000&source=images&cd=vfe&ved=0CAsQjRxqFwoTCKi6vZCf5fECFQAAAAAdAAAAABAP

Philosophy/Framework

Create lower time frame systems/strategies that excel in certain market conditions

Predict the likely regime/market characteristics over some larger time frame

Select the strategy most likely to be profitable in predicted regime

Price Components

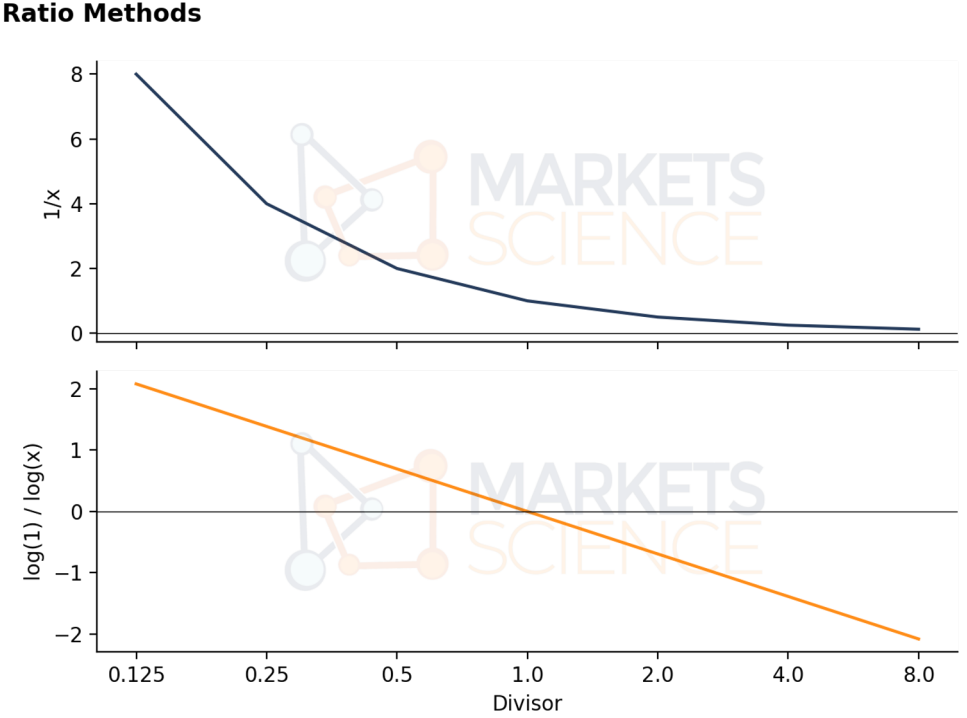
- Price movement can be broadly described by:
  - Direction
  - Volatility/Range
  - Trend
- These are the things we should be trying to predict

Useful Transformations

Log Ratio

$$ln(x_1) - ln(x_2)$$

```
def calc_log_ratio(s1, s2, eps=0):
    return np.log(s1 + eps) - np.log(s2 + eps)
```

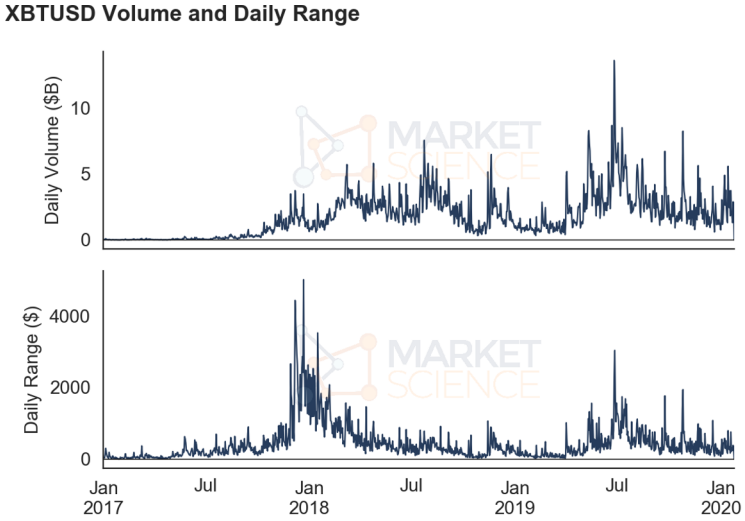


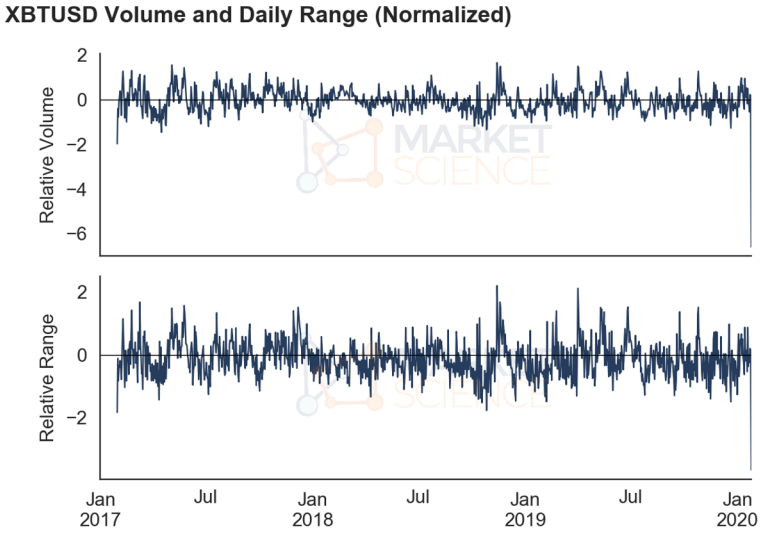
Useful Transformations

Relative Indicators

$$log(x) - log(\bar{x})$$

```
def relative_series_log(series, lookback=20, eps=0):
    s1 = series
    s2 = series.rolling(lookback, min_periods=2).mean().shift(1)
    return calc_log_ratio(s1, s2, eps)
```





Target Variables - Price Direction

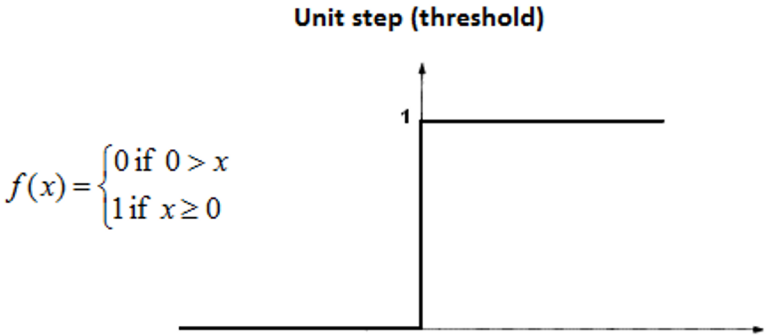
Log Return

$$log(x_{t+1}) - log(x_t)$$

```
def calc_log_return(ohlc):
    x1 = ohlc['open'].shift(-1)
    x2 = ohlc['open'].shift(-2)
    log_return = calc_log_ratio(x1, x2)

    return log_return
```

Price Direction



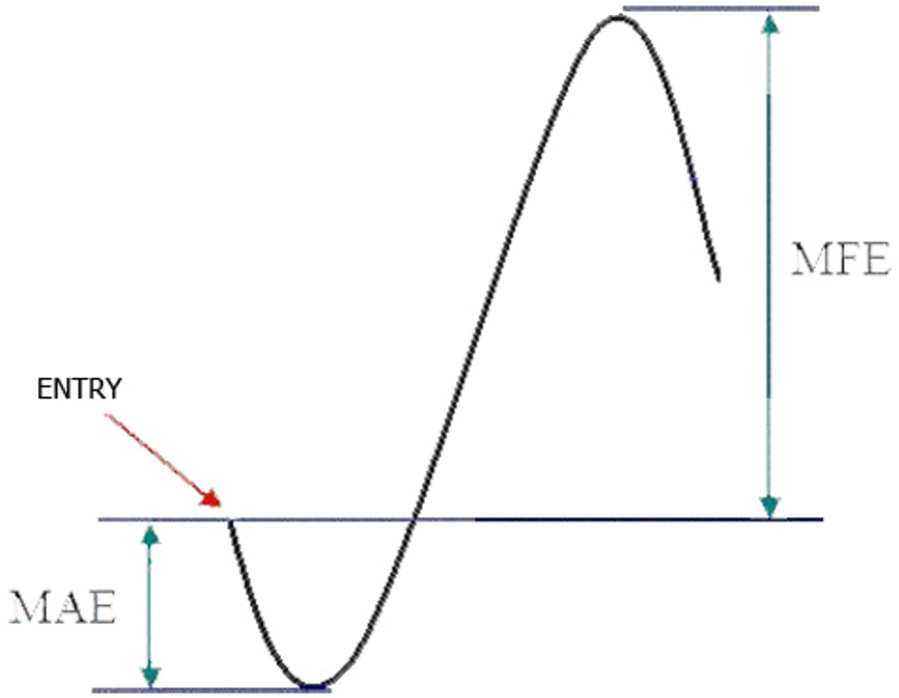
Target Variables - Price Direction

Maximum Positive Excursion

$$max(X) - x_t$$

```
def calc_mpe(ohlc, lookforward=0):
    max_high = (
        ohlc["high"]
        .rolling(lookforward + 1)
        .max()
        .shift(-lookforward)
    )
    mpe = max_high - ohlc["open"]

    return mpe
```



http://www.traderslaboratory.com/forums/topic/4537-analysing-quality-entry-with-e-ratio-maemfe/

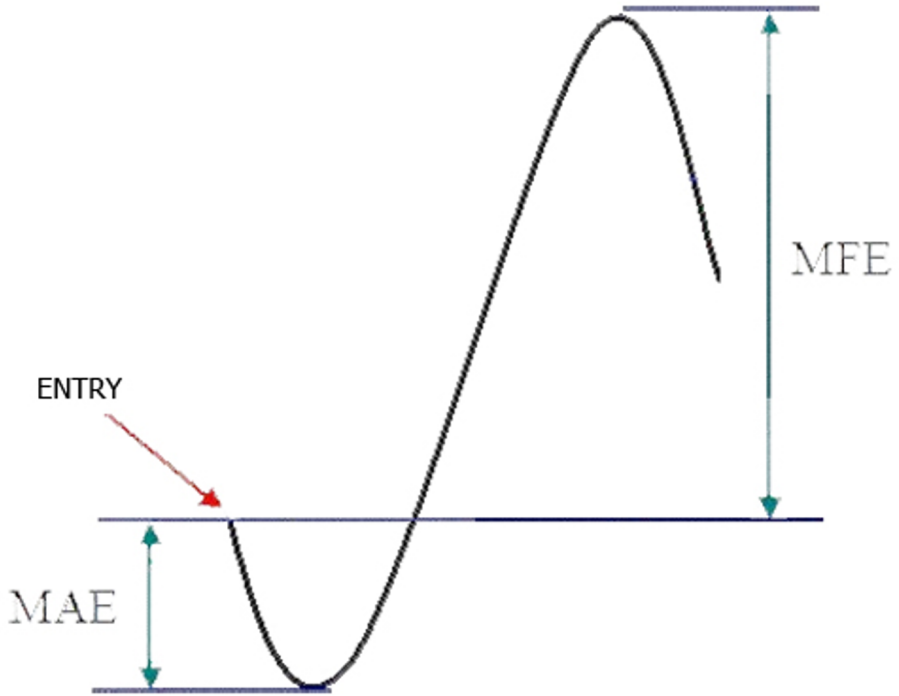
Target Variables - Price Direction

Maximum Negative Excursion

$$x_t - min(X)$$

```
def calc_mne(ohlc, lookforward=0):
    min_low = (
        ohlc["low"]
        .rolling(lookforward + 1)
        .min()
        .shift(-lookforward)
    )
    mne = ohlc["open"] - min_low

    return mne
```



http://www.traderslaboratory.com/forums/topic/4537-analysing-quality-entry-with-e-ratio-maemfe/

Target Variables - Price Direction

Edge Ratio (credit BuildAlpha)

$$\frac{log(MFE)}{log(MAE)}$$

```
def calc_edge_ratio_log(ohlc, lookforward=1, eps=1e-6, max_val=25):
    mpe = calc_mpe(ohlc, lookforward)
    mne = calc_mne(ohlc, lookforward)
    edge_ratio_log = calc_log_ratio(mpe, mne, eps=eps).clip(-max_val, max_val)

    return edge_ratio_log
```

PREFERRED METHOD

Target Variables - Range/Volatility

(Average) True Range (Percent)

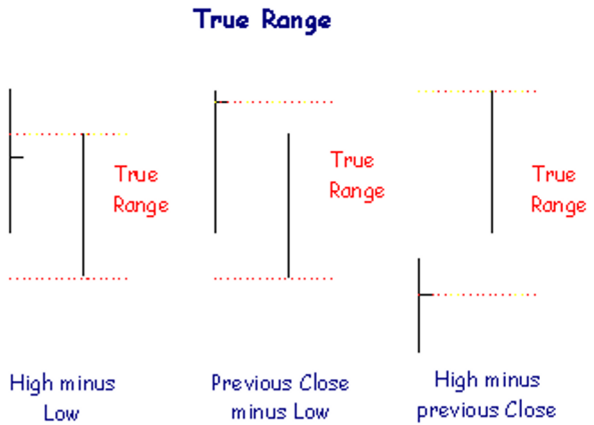
```python
def calc_true_range(ohlc):
    h, l, c = split_ohlc(ohlc)[1:4]
    method_1 = h - l
    method_2 = (h - c.shift(1)).abs()
    method_3 = (l - c.shift(1)).abs()
    true_range = pd.concat((method_1, method_2, method_3), axis=1).max(axis=1)

    return true_range


def calc_average_true_range(ohlc, lookback):
    true_range = calc_true_range(ohlc)
    average_true_range = true_range.rolling(lookback).mean()

    return average_true_range


def ATR_perc(ohlc, lookback):
    true_range = TR(ohlc)
    true_range_percent = true_range / ohlc["open"]
    average_true_range_percent = true_range_percent.rolling(lookback).mean()
    return average_true_range_percent
```



https://www.incrediblecharts.com/indicators/true_range.php

Target Variables - Range/Volatility

Tape Length

$$\sum |close_{LTF} - open_{LTF}|$$

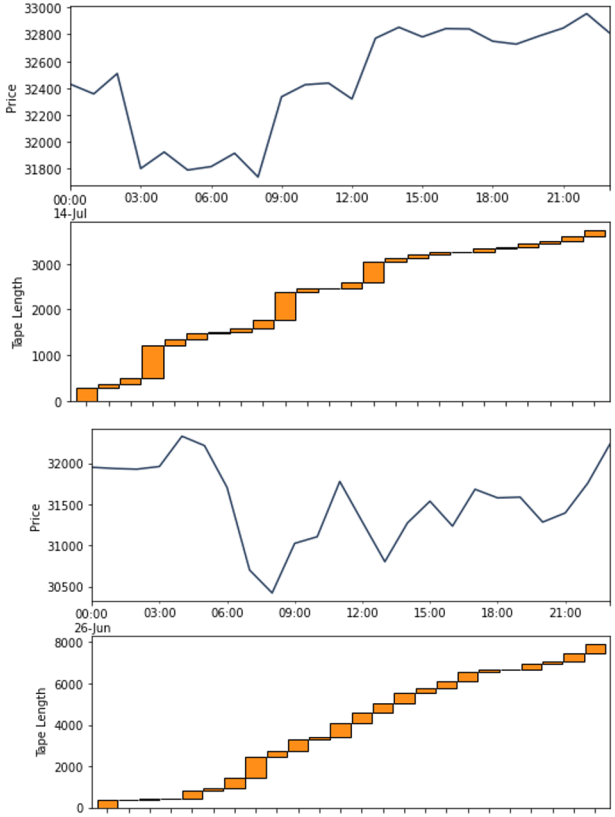
```
def calc_tape_length(price_series):
    abs_diffs = price_series.diff().abs()
    tape_length = abs_diffs.sum()

    return tape_length
```

Total Move

```
def calc_total_move(price_series):
    total_move = abs(price_series[-1] - price_series[0])

    return total_move
```



Target Variables - Trend

Efficiency Ratio

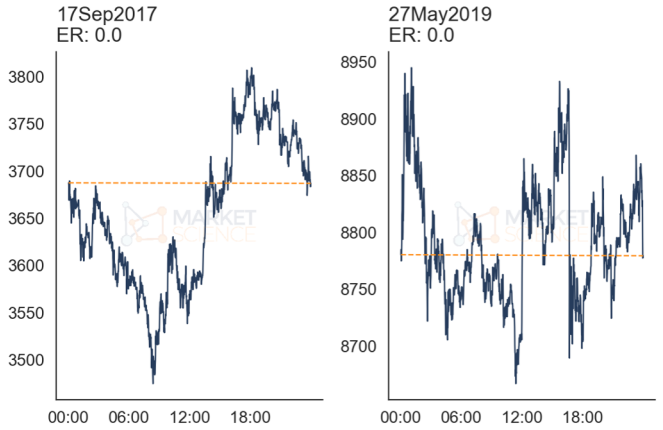
*credit: Perry Kaufman

- Compares total move to the amount of intra-period movement
- Tape length is the denominator

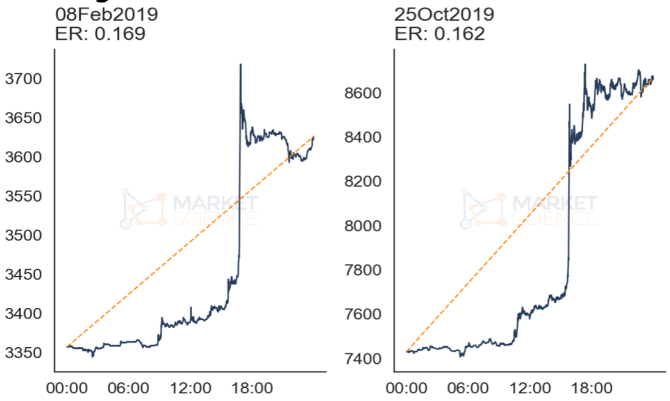
$$\frac{|close_{HTF} - open_{HTF}|}{\sum |close_{LTF} - open_{LTF}|}$$

```
def calc_efficiency_ratio(series):
    total_diff = abs(series[-1] - series[0])
    sum_diffs = series.diff().abs().sum()
    return total_diff / sum_diffs
```

Weak



Strong



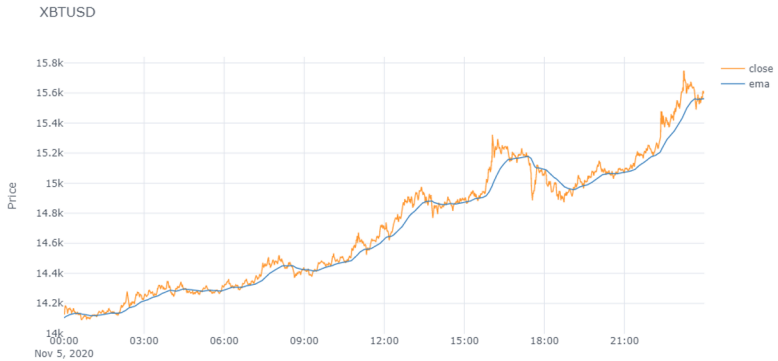
Target Variables - Trend

Moving Average Dominance

- Calculates the proportion of time price spends on one side of a moving average
- Trending markets cross their MAs more rarely

```
def calc_moving_average_dominance(series, lookback):
    ema = series.ewm(span=lookback).mean()
    over_ema = series > ema
    mad_raw = over_ema.sum() / len(series)
    moving_average_dominance = max(mad_raw, 1-mad_raw)

    return moving_average_dominance
```



Model Creation

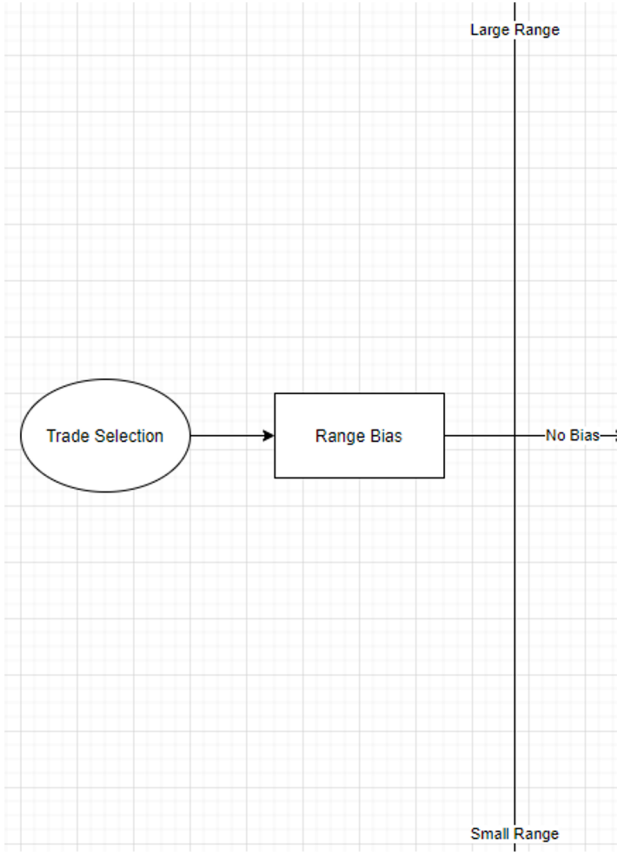
Feature Selection

- For each target variable, determine which market features are most important
- Will most likely use different features for each target

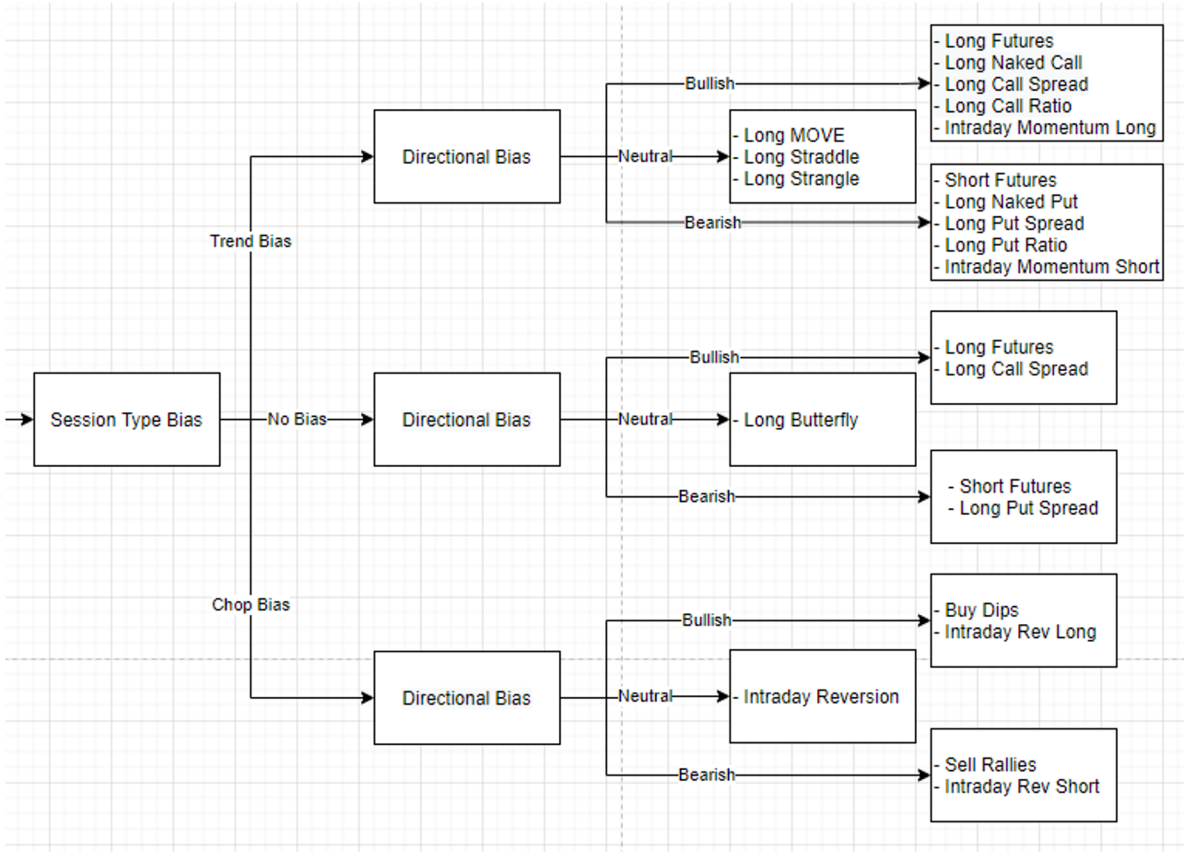
Inputs

- Seasonality/Time-of-Day effects
- Statistical Anomalies in price action
- CME positioning
- Sentiment (social media)
- Token flows
  - To/from exchanges
  - Stablecoin creation at treasuries
- Rates (futures curve premiums/perp swaps funding rates)

Trade Selection - Range Bias

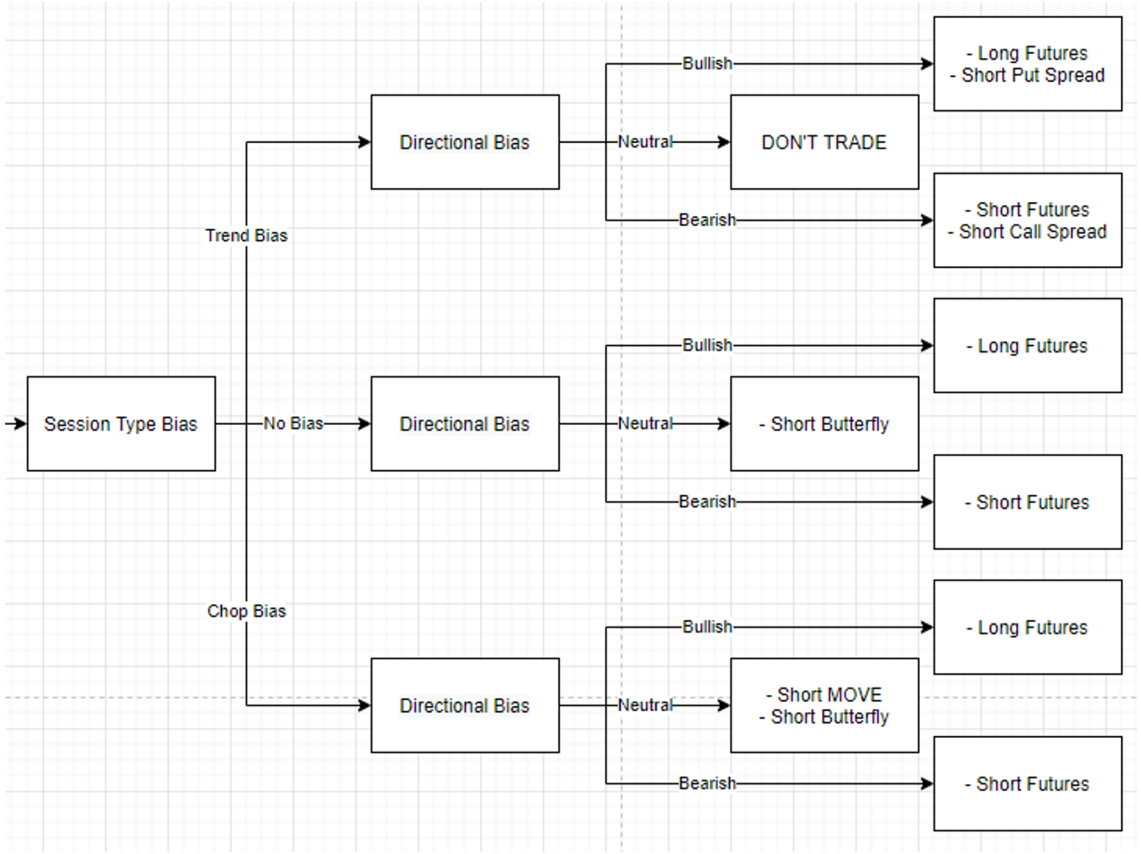


Trade Selection - Large Range



Trade Selection - No Range Bias

Trade Selection - Small Range



Contact

Linkedin: Brian Blandin

Twitter: @quantfiction

Quantfiction.com

marketsscience.com