### Key Steps in Pairs Trading: Detailed Statistical Approach

Here's a comprehensive guide to executing a pairs trading strategy with a focus on statistical concepts, using the WTI Crude Oil (CL) and Heating Oil (HO) pair as an example.

#### 1. Selection of Pairs

\*\*Rationale:\*\*

- WTI and Heating Oil have a logical economic relationship because Heating Oil is a refined product of crude oil.

- Historical price relationships can be exploited for mean reversion opportunities.

#### 2. Data Collection

\*\*Requirements:\*\*

- \*\*Historical Prices:\*\* Obtain daily closing prices for WTI Crude Oil and Heating Oil futures. You may need data spanning several years for reliable statistical analysis.

- \*\*High-Frequency Data:\*\* If intraday trading, gather high-frequency data.

#### 3. Data Preparation

\*\*Steps:\*\*

- \*\*Data Cleaning:\*\* Ensure data is free from errors, missing values, and inconsistencies.

- \*\*Normalization:\*\* Adjust for splits, dividends, and other corporate actions if dealing with stocks, not applicable directly here.

#### 4. Statistical Analysis

\*\*Key Concepts:\*\*

##### Stationarity

\*\*Definition:\*\*

- A stationary time series has a constant mean, variance, and autocorrelation structure over time.

\*\*Tests:\*\*

- \*\*Augmented Dickey-Fuller (ADF) Test:\*\* Checks for unit roots.

- \*\*KPSS Test:\*\* Tests for stationarity around a mean or trend.

\*\*Procedure:\*\*

1. \*\*ADF Test:\*\* Apply to both WTI and Heating Oil prices.

```python

from statsmodels.tsa.stattools import adfuller

result = adfuller(price\_series)

print(f'ADF Statistic: {result[0]}')

print(f'p-value: {result[1]}')

```

2. \*\*KPSS Test:\*\*

```python

from statsmodels.tsa.stattools import kpss

result = kpss(price\_series)

print(f'KPSS Statistic: {result[0]}')

print(f'p-value: {result[1]}')

```

##### Cointegration

\*\*Definition:\*\*

- Two or more non-stationary series are cointegrated if some linear combination of them is stationary.

\*\*Tests:\*\*

- \*\*Engle-Granger Test:\*\* Residual-based test for pairwise cointegration.

- \*\*Johansen Test:\*\* Tests for multiple cointegrating vectors.

\*\*Procedure:\*\*

1. \*\*Engle-Granger Test:\*\*

- Run a linear regression between WTI and Heating Oil.

```python

import statsmodels.api as sm

X = sm.add\_constant(WTI\_prices)

model = sm.OLS(HO\_prices, X).fit()

residuals = model.resid

result = adfuller(residuals)

print(f'ADF Statistic: {result[0]}')

print(f'p-value: {result[1]}')

```

- Check if residuals are stationary (low p-value in ADF test).

2. \*\*Johansen Test:\*\*

```python

from statsmodels.tsa.vector\_ar.vecm import coint\_johansen

result = coint\_johansen(data, det\_order, k\_ar\_diff)

print(f'Trace Statistic: {result.lr1}')

print(f'Critical Values: {result.cvt}')

```

##### Linear Regression

\*\*Definition:\*\*

- Linear regression models the relationship between a dependent variable and one or more independent variables.

\*\*Procedure:\*\*

1. \*\*Fit Regression Model:\*\*

```python

X = sm.add\_constant(WTI\_prices)

model = sm.OLS(HO\_prices, X).fit()

print(model.summary())

```

2. \*\*Interpret Coefficients:\*\* Assess the slope and intercept to understand the relationship dynamics.

3. \*\*Residual Analysis:\*\* Ensure residuals are white noise (normally distributed with constant variance).

##### Spread Calculation and Analysis

\*\*Steps:\*\*

1. \*\*Calculate Spread:\*\* Define spread as the residuals from the cointegration regression.

```python

spread = HO\_prices - model.predict(X)

```

2. \*\*Plot Spread:\*\* Visualize to check for mean-reverting behavior.

```python

spread.plot()

```

#### 5. Entry and Exit Strategies

\*\*Signal Generation:\*\*

- \*\*Thresholds:\*\* Use statistical measures (e.g., standard deviations) to set entry/exit points.

```python

mean\_spread = spread.mean()

std\_spread = spread.std()

long\_entry = mean\_spread - 2 \* std\_spread

short\_entry = mean\_spread + 2 \* std\_spread

```

\*\*Trade Execution:\*\*

- \*\*Entry:\*\* Enter long (WTI) and short (HO) when the spread is below the lower threshold.

- \*\*Exit:\*\* Close positions when the spread reverts to the mean.

#### 6. Position Sizing

\*\*Risk Management:\*\*

- \*\*Volatility-Based Sizing:\*\* Adjust position sizes based on the volatility of the spread.

```python

position\_size = capital \* (1 / std\_spread)

```

#### 7. Performance Measurement

\*\*Metrics:\*\*

- \*\*Sharpe Ratio:\*\* Measures risk-adjusted return.

- \*\*Maximum Drawdown:\*\* Assesses potential loss.

- \*\*Win/Loss Ratio:\*\* Tracks trade success rate.

\*\*Backtesting:\*\*

- \*\*Historical Simulation:\*\* Apply strategy on historical data.

- \*\*Out-of-Sample Testing:\*\* Validate on unseen data.

### Conclusion

By following these key steps and focusing on statistical concepts like stationarity, cointegration, and linear regression, you can effectively implement a pairs trading strategy with WTI Crude Oil and Heating Oil. This approach requires rigorous statistical analysis and continuous monitoring to adapt to changing market conditions.