INTRODUCTION TO TIME SERIES ANALYSIS IN PYTHON

Cointegration Models

the idea behind cointegration is that even if the prices of two different assets both follow random walks, it is still possible that a linear combination of them is not a random walk

Rob Reider
Adjunct Professor, NYU-Courant
Consultant, Quantopian

What is Cointegration?

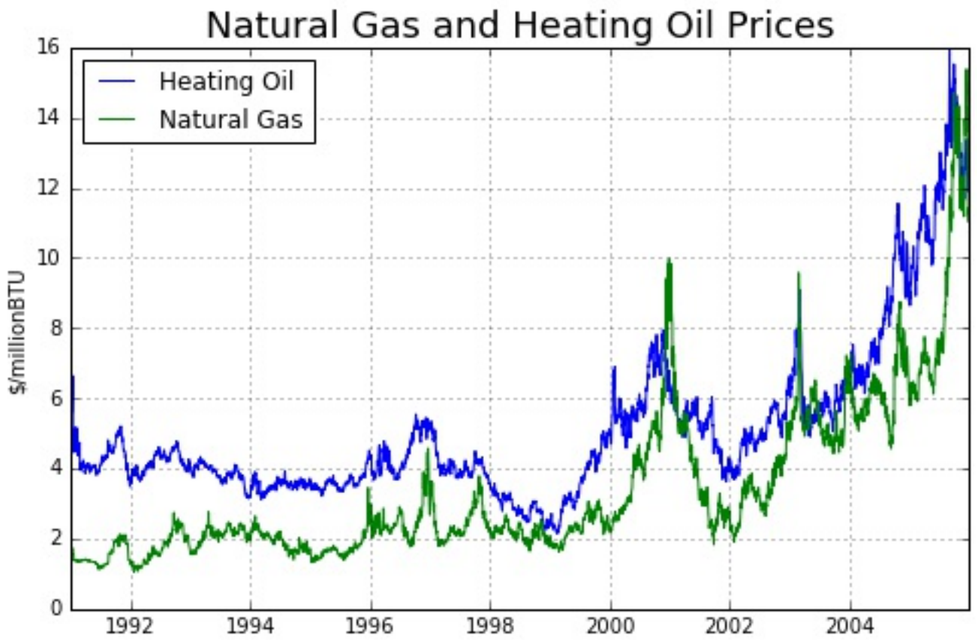
- Two series, P_t and Q_t can be random walks
- But the linear combination $P_t - c\,Q_t$ may not be a random walk!
- If that's true (P and Q are not forecastable bc they are random walks)
  - $P_t - c\,Q_t$ is forecastable
  - P_t and Q_t are said to be cointegrated

Analogy: Dog on a Leash

- $P_t =$ Owner
- $Q_t =$ Dog
- Both series look like a random walk
- Difference, or distance betweem them, looks mean reverting
  - If dog falls too far behind, it gets pulled forward
  - If dog gets too far ahead, it gets pulled back

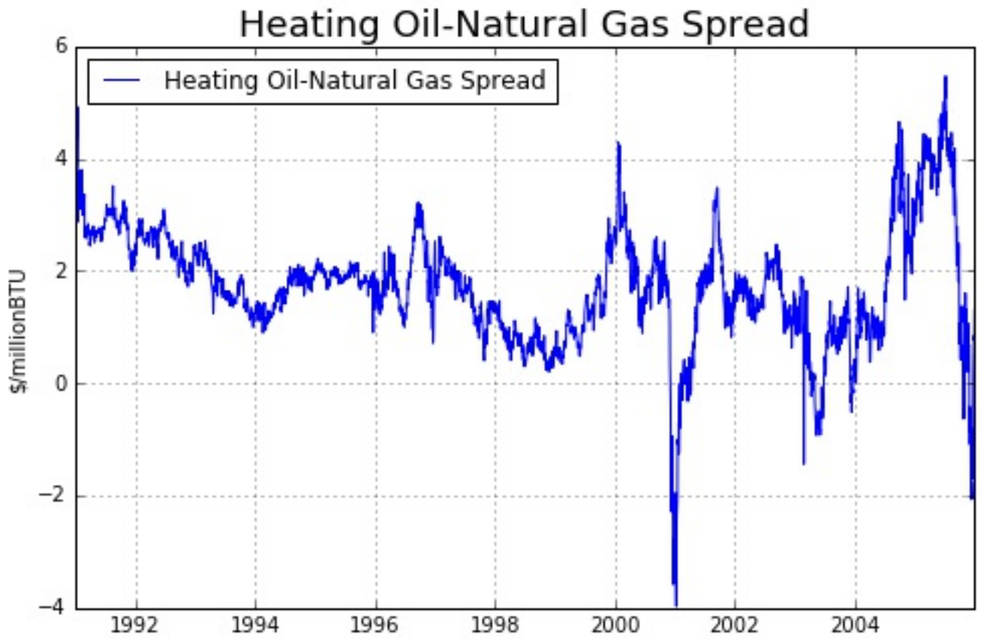
Example: Heating Oil and Natural Gas

- Heating Oil and Natural Gas both look like random walks...



Example: Heating Oil and Natural Gas

- But the spread (difference) is mean reverting



What Types of Series are Cointegrated?

- Economic substitutes
  - Heating Oil and Natural Gas
  - Platinum and Paladium
  - Corn and Wheat
  - Corn and Sugar
  - ...
  - Bitcoin and Ethereum?
- How about competitors?
  - Coke and Pepsi?
  - Apple and Blackberry? No! Leash broke and dog ran away

with commodities, there may be economic forces that link the two prices

for stocks, a natural starting point for identifying cointegrated pairs are stocks in the same industry

competitors are not necessarily economic substitutes

Two Steps to Test for Cointegration

process for testing whether two series are cointegrated into 2 steps

- Regress P_t on Q_t and get slope c
- Run Augmented Dickey-Fuller test on $P_t - c\,Q_t$ to test for random walk
- Alternatively, can use coint function in statsmodels that combines both steps

```
from statsmodels.tsa.stattools import coint
coint(P,Q)
```

INTRODUCTION TO TIME SERIES ANALYSIS IN PYTHON

Let's practice!

INTRODUCTION TO TIME SERIES ANALYSIS IN PYTHON

Case Study: Climate Change

Rob Reider
Adjunct Professor, NYU-Courant
Consultant, Quantopian

Analyzing Temperature Data

- Temperature data:
  - New York City from 1870-1916
  - Downloaded from National Oceanic and Atmospheric Administration (NOAA)
- Convert index to datetime object
- Plot data

Analyzing Temperature Data

- Test for Random Walk
- Take first differences
- Compute ACF and PACF
- Fit a few AR, MA, and ARMA models
- Use Information Criterion to choose best model
- Forecast temperature over next 30 years

INTRODUCTION TO TIME SERIES ANALYSIS IN PYTHON

Let's practice!

INTRODUCTION TO TIME SERIES ANALYSIS IN PYTHON

Congratulations

Rob Reider
Adjunct Professor, NYU-Courant
Consultant, Quantopian

Advanced Topics

- GARCH Models
- Nonlinear Models
- Multivariate Time Series Models
- Regime Switching Models
- State Space Models and Kalman Filtering
- ...

INTRODUCTION TO TIME SERIES ANALYSIS IN PYTHON

Keep practicing!