

# Cointegration Models

TIME SERIES ANALYSIS IN PYTHON



**Rob Reider**

Adjunct Professor, NYU-Courant  
Consultant, Quantopian

# What is Cointegration?

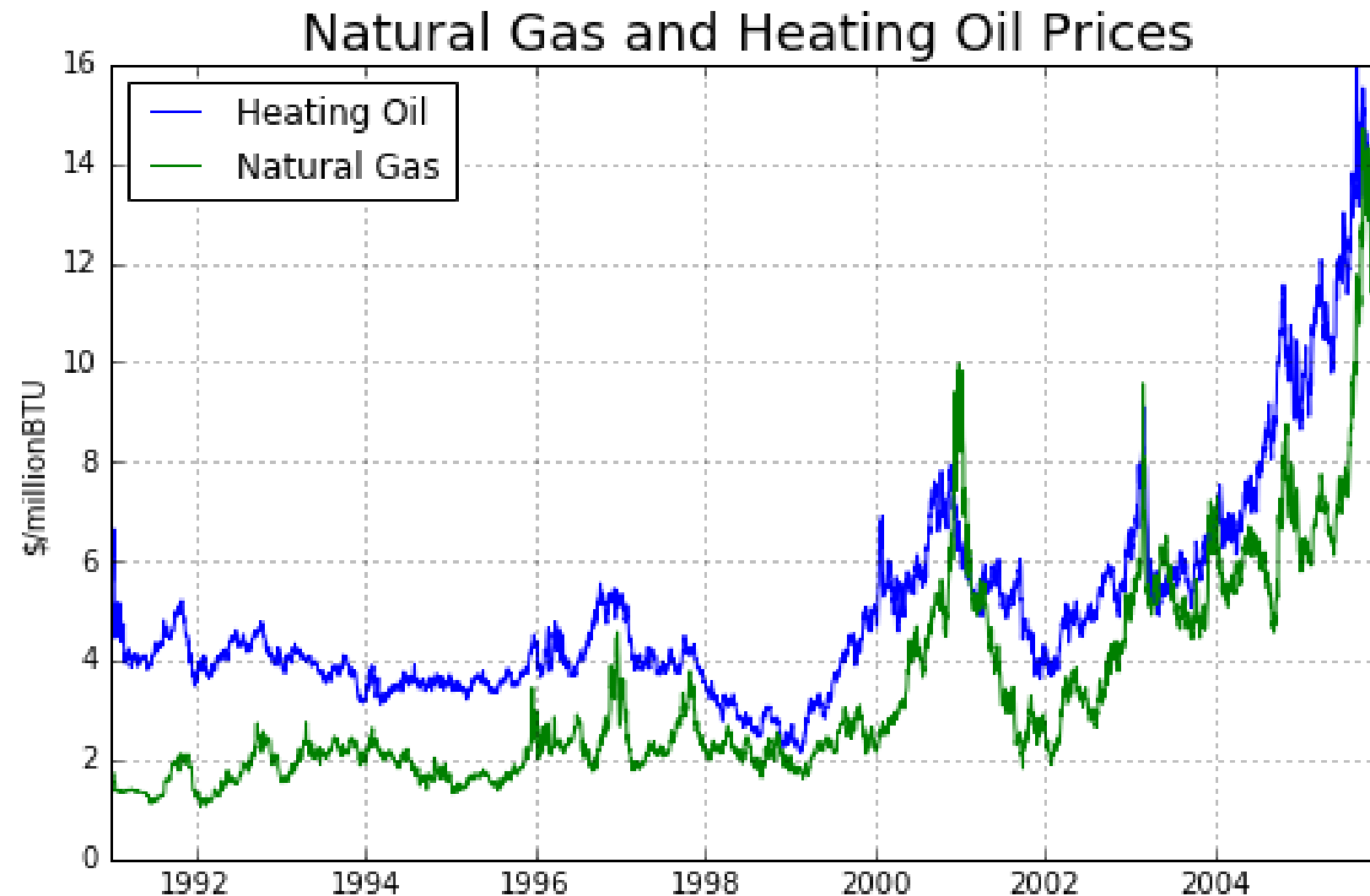
- Two series,  $P_t$  and  $Q_t$  can be random walks
- But the linear combination  $P_t - c Q_t$  may not be a random walk!
- If that's true
  - $P_t - c Q_t$  is forecastable
  - $P_t$  and  $Q_t$  are said to be cointegrated

# Analogy: Dog on a Leash

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

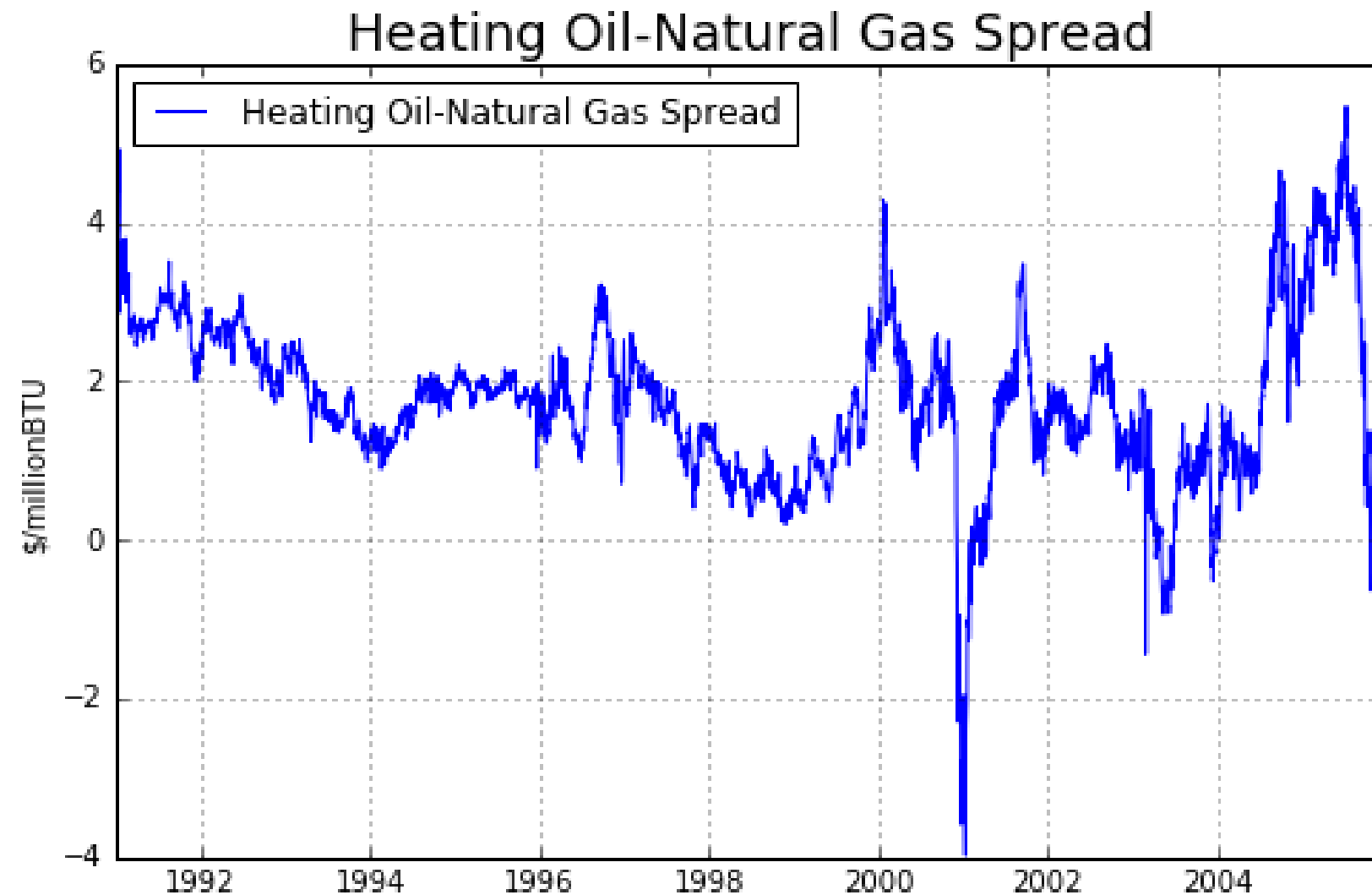
# Example: Heating Oil and Natural Gas

- 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 Palladium
  - 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

# Two Steps to Test for Cointegration

- 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)
```

# Let's practice!

TIME SERIES ANALYSIS IN PYTHON



# Case Study: Climate Change

TIME SERIES ANALYSIS IN PYTHON



**Rob Reider**

Adjunct Professor, NYU-Courant  
Consultant, Quantopian

# Analyzing Temperature Data

- Temperature data:
  - New York City from 1870-2016
  - 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

# Let's practice!

TIME SERIES ANALYSIS IN PYTHON

# Congratulations

TIME SERIES ANALYSIS IN PYTHON



**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
- ...

**Keep practicing!**  
TIME SERIES ANALYSIS IN PYTHON