Machine Learning and Finance: Through the Ages

Tim Savage

Senior Managing Economist

Principal Data Scientist
CBRE

Usual Disclaimer

These are my opinions, and not those of CBRE.

Or anyone else.

Some Motivating Examples

Prediction

- Based a training set of labeled objects, what is this unlabeled object?
- Ideally using validation and tuning.
- Always probabilistic: the unlabeled is a zero with probability 98.2%.

Evaluation

- What is the effect of this drug?
- Randomize the unrandomized over a treatment.
- Double-blind experimentation: observation does not affect outcome.
- Ideally probabilistic (but almost always is not).

Similarity (Absent a Biological Model)

- Given some massive genome data, are there “disease” clusters?

This Talk

- Given my risk appetite and cash constraints, can I make money?

Two Images: You Tell Me

Image One

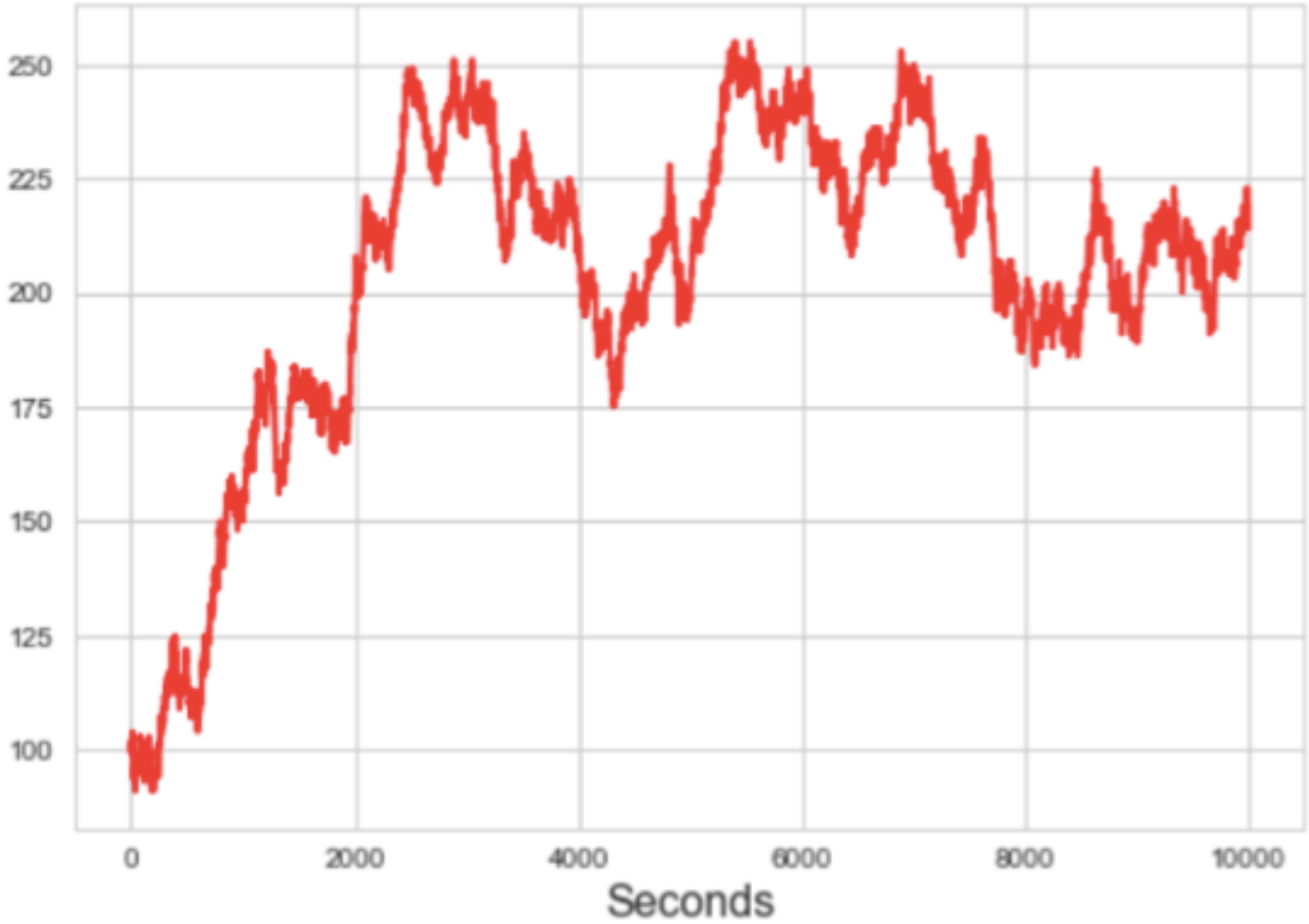

250
225
200
175
150
125
100
0
2000
4000
6000
8000
10000
Seconds

Image Two

70
65
60
55
50
45
40
$
2012-02
2012-03
2012-04
2012-05
2012-06
2012-07
2012-08
2012-09
2012-10
2012-11
2012-12
Date

Image One: A One-Dimensional Random Walk

```
random.seed(1234)

pos = 100
walk = [pos] # walk is the array that tracks the random path
nsteps = 10000 # establishs number of random steps
for i in range(nsteps):
    step = 1 if np.random.randint(0, 2) else -1 # Bernoulli draw to step "up" or "down"
    pos += step
    walk.append(pos)
```

Image Two: AAPL (2012)

```
start, end = "2012-01-01", "2013-01-01"
aapl = web.get_data_yahoo('aapl', start=start, end=end)['Adj Close']

aapl.plot(color = 'darkblue')
plt.title('Apple Share Price (2012)', fontsize=20)
plt.ylabel('$', fontsize=16)
```

Some Background on Modern Finance

- Bachelier (1900) proposes that stock prices are a random walk.
  - Not quite because prices cannot be negative.
  - But changes in prices can be positive or negative.
- As equity markets developed, individual stocks traded on exchanges.
  - Think NASDAQ and tech stocks.
- Emergence of portfolios that include equities, bonds, and cash.
- Diversification is key.
  - Reduce risk for a given level of positive return (positive price changes).

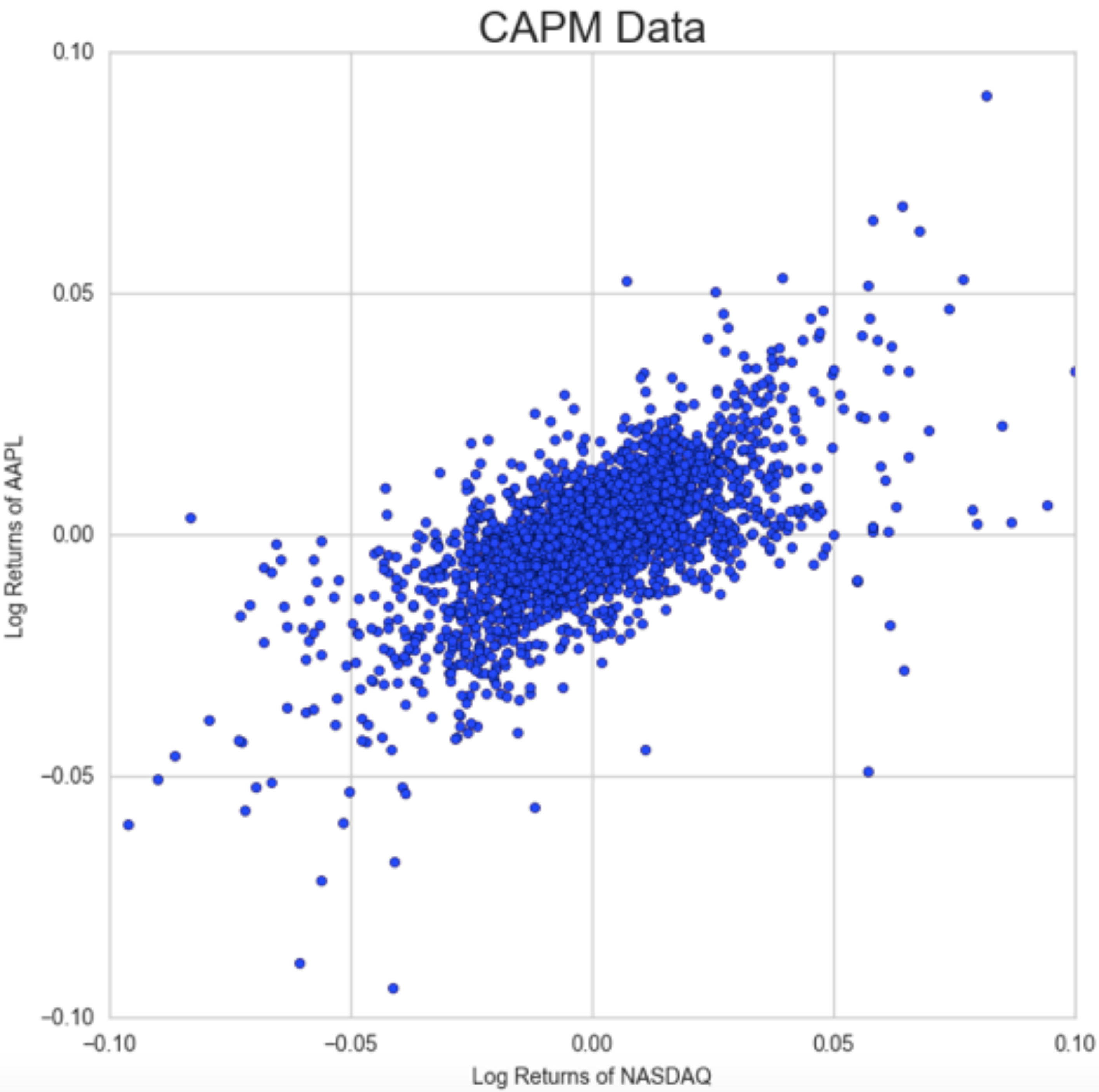
Interesting Use Case for Machine Learning

- But data science requires, among other things, …
  - Real-world data.
- Data on prices are among the oldest data series that exist.
- Originally, very low frequency.
- Now tick data are generated on the range of a microsecond.
  - Faster than you can click ”Like”.

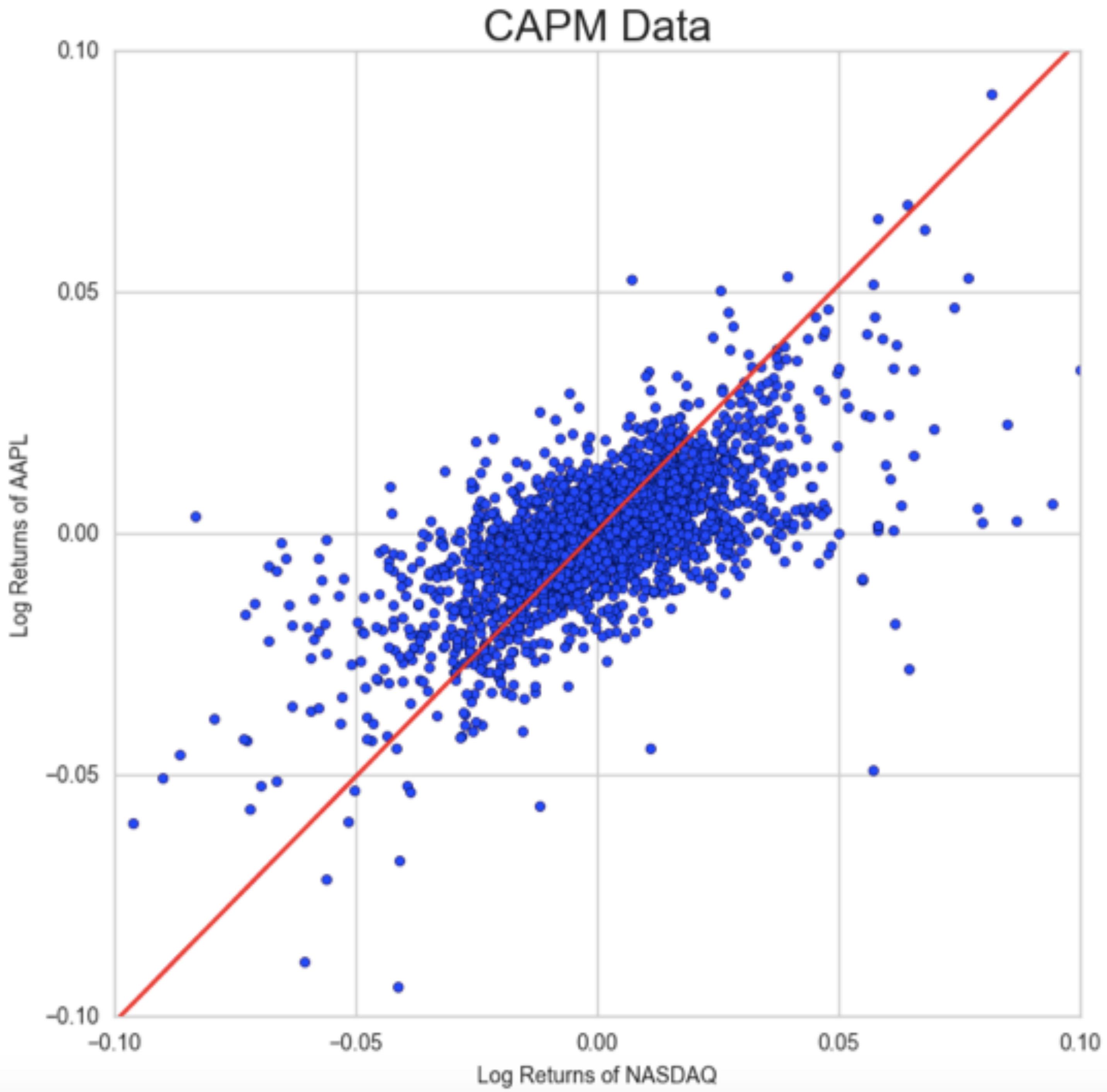
Wave One:
Theory and Testing


AAPL ($/Share)
140
120
100
80
60
40
20
0
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
Date


NASDAQ
5500
5000
4500
4000
3500
3000
2500
2000
1500
1000
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
Date



CAPM Data
0.10
0.05
0.00
−0.05
−0.10
Log Returns of AAPL
−0.10
−0.05
0.00
0.05
0.10
Log Returns of NASDAQ


CAPM Data
0.10
0.05
0.00
−0.05
−0.10
Log Returns of AAPL
−0.10
−0.05
0.00
0.05
0.10
Log Returns of NASDAQ

Arbitrage Portfolio Theory (APT)

- Costlessly adjust components of portfolios, relative to a benchmark.
- If true, linear regression may be an ideal representation.
- Slope captures non-diversifiable risk, given the benchmark.
  - Greater or less than one?
- Bias term captures ”talent” of active management, given non-diversifiable risk.
  - Greater or less than zero?

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                   aapl   R-squared:                       0.433
Model:                            OLS   Adj. R-squared:                  0.433
Method:                 Least Squares   F-statistic:                     2114.
Date:                Wed, 12 Apr 2017   Prob (F-statistic):               0.00
Time:                        20:05:56   Log-Likelihood:                 7541.1
No. Observations:                2768   AIC:                        -1.508e+04
Df Residuals:                    2766   BIC:                        -1.507e+04
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [95.0% Conf. Int.]
------------------------------------------------------------------------------
Intercept      0.0006      0.000      1.898      0.058     -1.89e-05     0.001
nasdaq         1.0184      0.022     45.975      0.000         0.975     1.062
==============================================================================
Omnibus:                      463.646   Durbin-Watson:                   1.933
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             8105.015
Skew:                           0.209   Prob(JB):                         0.00
Kurtosis:                      11.373   Cond. No.                         73.4
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

Wave Two: Prediction

Surely Deep Learning Will Improve on Regression

Give the Machines a Break!

Table 1: CAPM with AAPL Daily Returns			
Model	**Run Time[1]**	**Average MSE[2]**	**Smallest MSE[2]**
OLS	0.009	0.264	0.209
MLP	8.557	0.274	0.217
LSTM[3]	12.170	0.436	0.355
LSTM[4]	19.064	0.384	0.304

1: Average seconds per replication
2: 10^{-3}
3: 30-day window
4: 90-day window

Source: Savage and Vo, 2017

Try Something Different: Deploy Bayes

Within This Framework, Examine Something Meaningful

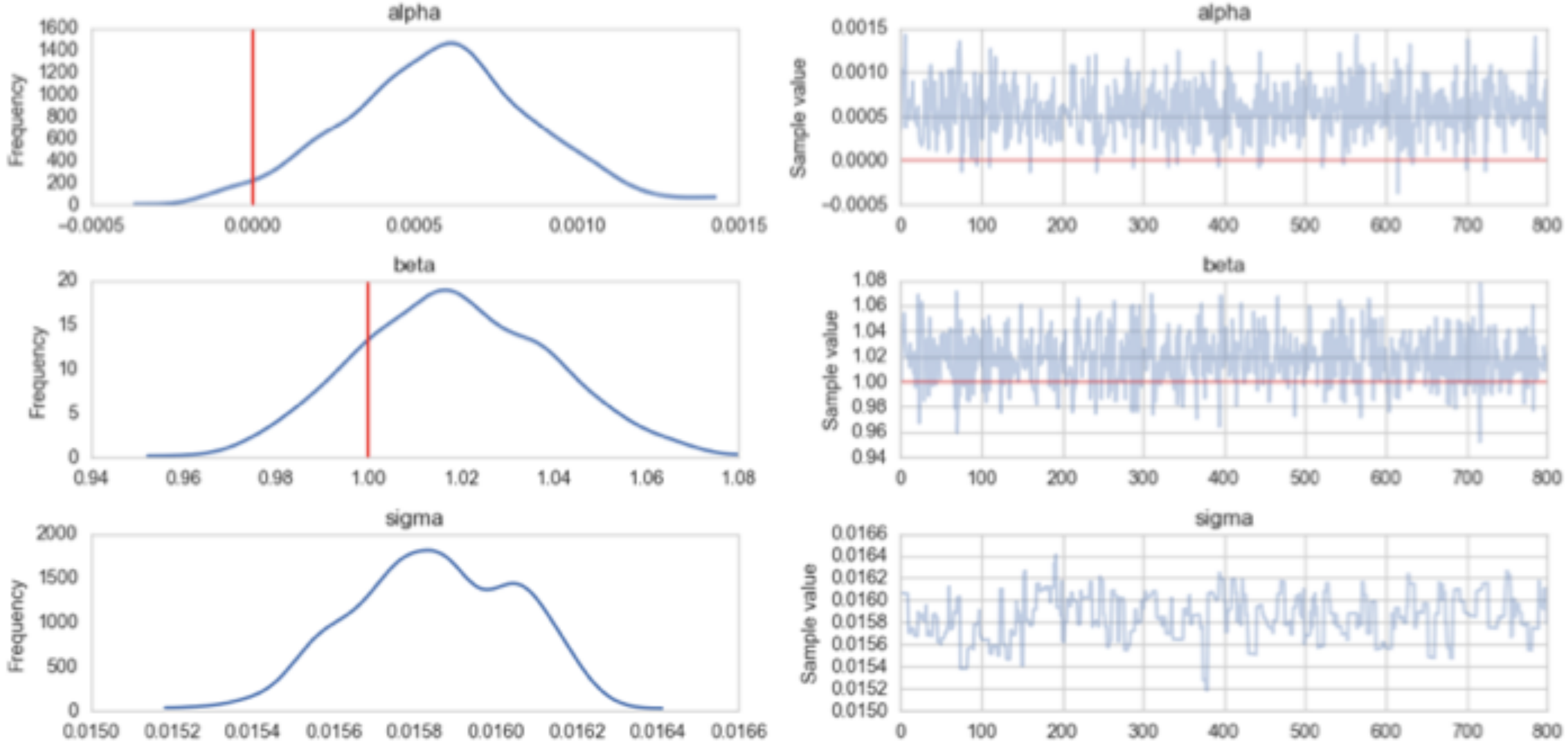
Some Code

```
with pm.Model() as model:
    # alpha, beta, and sigma are the hyperparameters over which we have our priors, in this case t
hey are flat priors.
    alpha = pm.Normal('alpha', mu=0, sd=20)
    beta = pm.Normal('beta', mu=0, sd=20)
    sigma = pm.Uniform('sigma', lower=0, upper=10)

    # y_est is the specification of the Bayesian model to be estimated.  It is simply our CAPM.
    y_est = alpha + beta * nasdaq_returns

    # likelihood is the likelihood function, here it is normal to be used with conjugate priors.
    likelihood = pm.Normal('y', mu=y_est, sd=sigma, observed=aapl_returns)

    # We use the Maximum a Posteriori (MAP) values as starting values for the MCMC sampling.
    start = pm.find_MAP()
    step = pm.NUTS(state=start)
    trace = pm.sample(1000, step, start=start, progressbar=True)
```

Some Results

```
# Show results after burn in of 200 MCMC replications.

fig = pm.traceplot(trace[199:1000], lines={'alpha': 0, 'beta': 1})
plt.figure(figsize = (10, 10))
```

```
<matplotlib.figure.Figure at 0x115bfbf60>
```



Evaluate Something Meaningful

- What is joint probability that alpha (bias term) is > 0 while beta (slope term) is <= 1?
- In this use case, Bayesian approach allows us to evaluate this directly.
  - Pr(alpha) > 0: 97.2%
  - Pr(beta) <= 1: 19.5%
  - Joint probability: 18.9%
- Implication?

Q&A