Finding Market Regimes in Equity Returns Using Hidden Markov Models

Nick Nelson - Computing Comps (Advisor M. Wade)

Tools & Resources

Software:

1 Python

1.1 Yfinance (stock data)

1.2 Numpy/Pandas (data manipulation)

1.3 Matplotlib (plotting)

1.4 HmmLearn (HMM)

2 Github (Version Control)

Hardware:

Personal Laptop

Advisor/ Consultant: Professor M. Wade & Professor M. Gee

About Me

BACKGROUND

Interest: Quantitative Finance and Machine Learning

Major: Math (Column D & E)

Relevant coursework: Mathematical Models, Optimal Control, Probability, Applied Linear Algebra, Stats (106 & 206)

School Experience: (Tutor/TA for COMP 118 & 218)

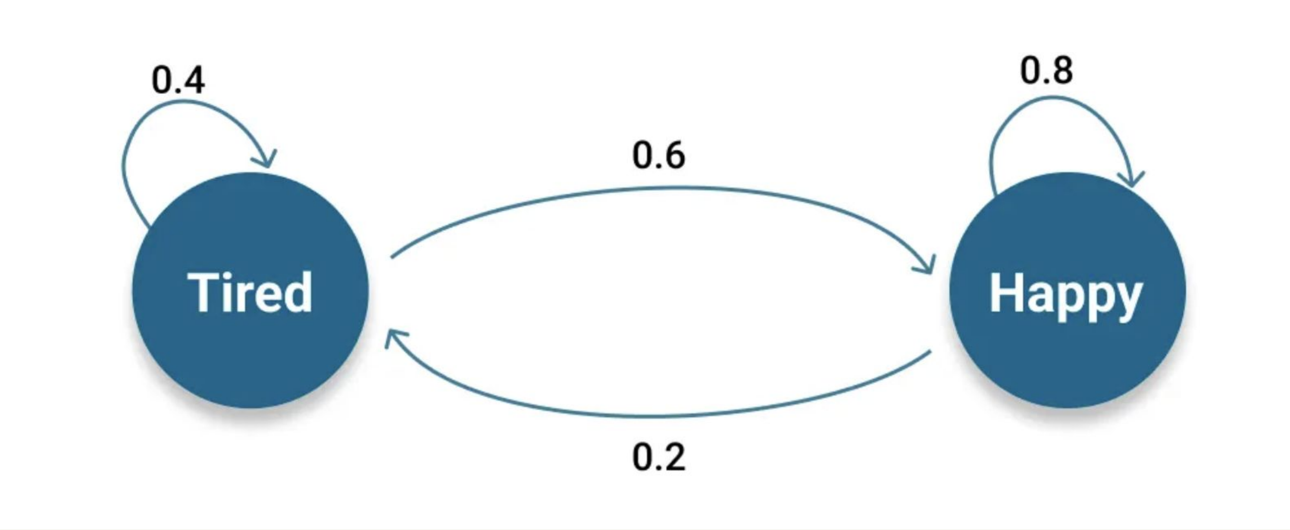
Professional Experience: 3 Data Science Internships; 2 of which were at JPMorgan Chase

Data Science → Quant Research

Research Question without Jargon

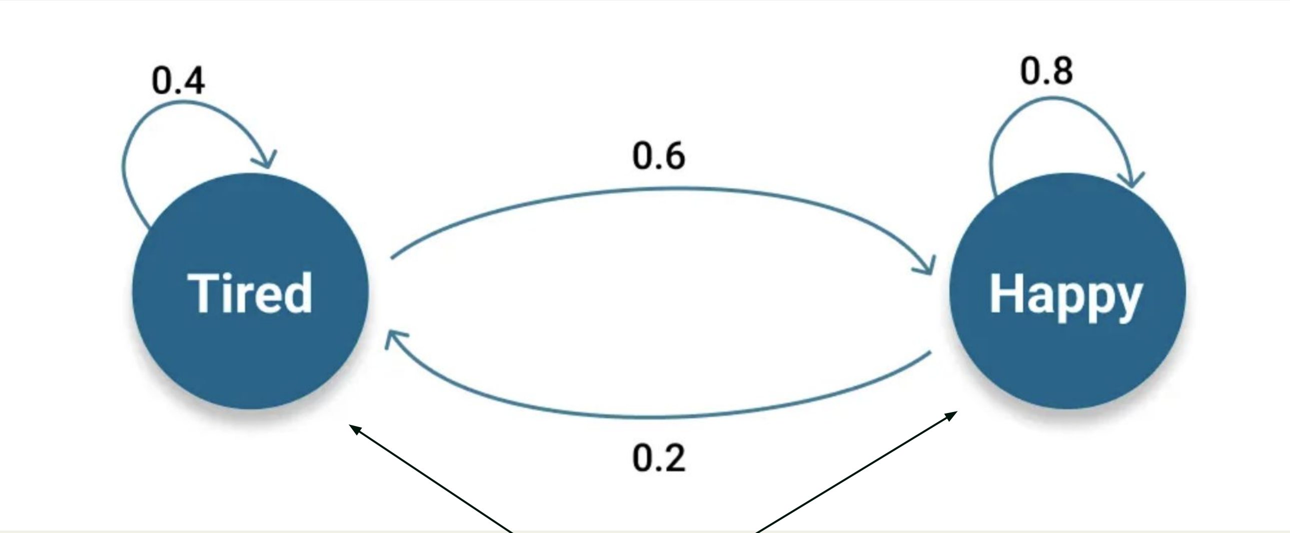
Can we automatically detect when the market is in a "calm" period versus a "volatile" period by only looking at daily price changes?"

Markov Chain



- stochastic process
- moves between states over time
- “Memoryless,” The next state only depends on the current state

Markov Chain

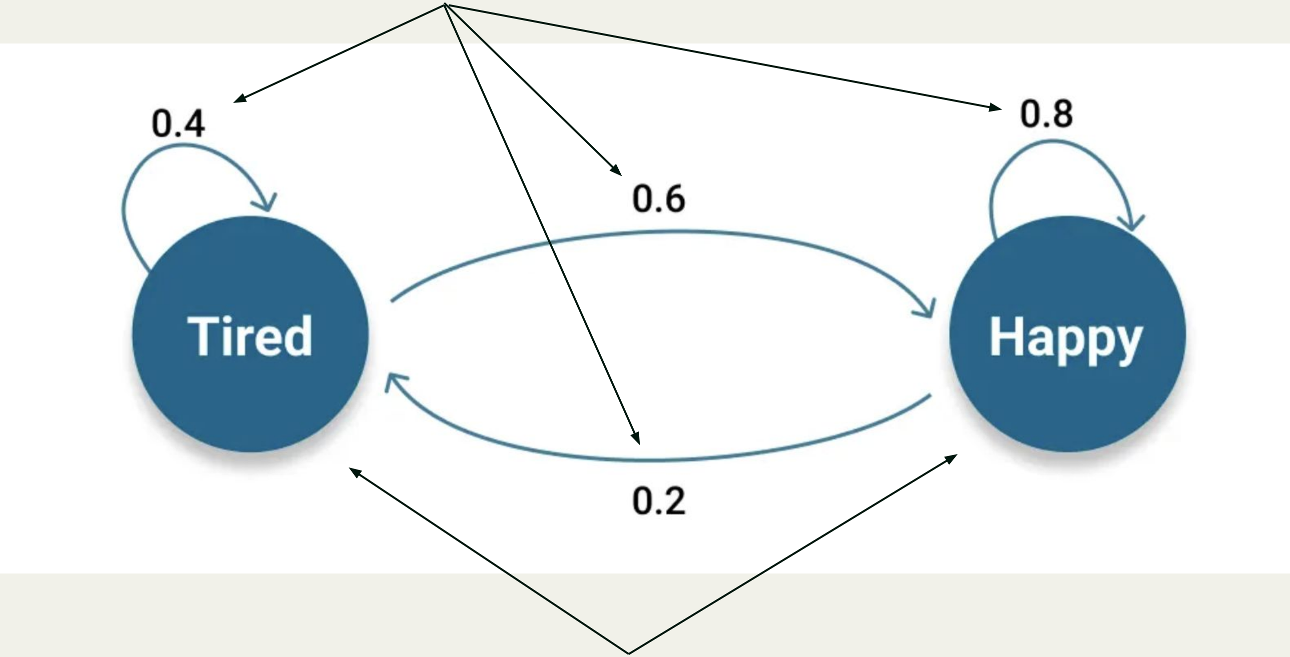


States: the environment/system

- stochastic process
- moves between states over time
- “Memoryless,” The next state only depends on the current state

Markov Chain

Transition Probabilities: from a current state to a next state

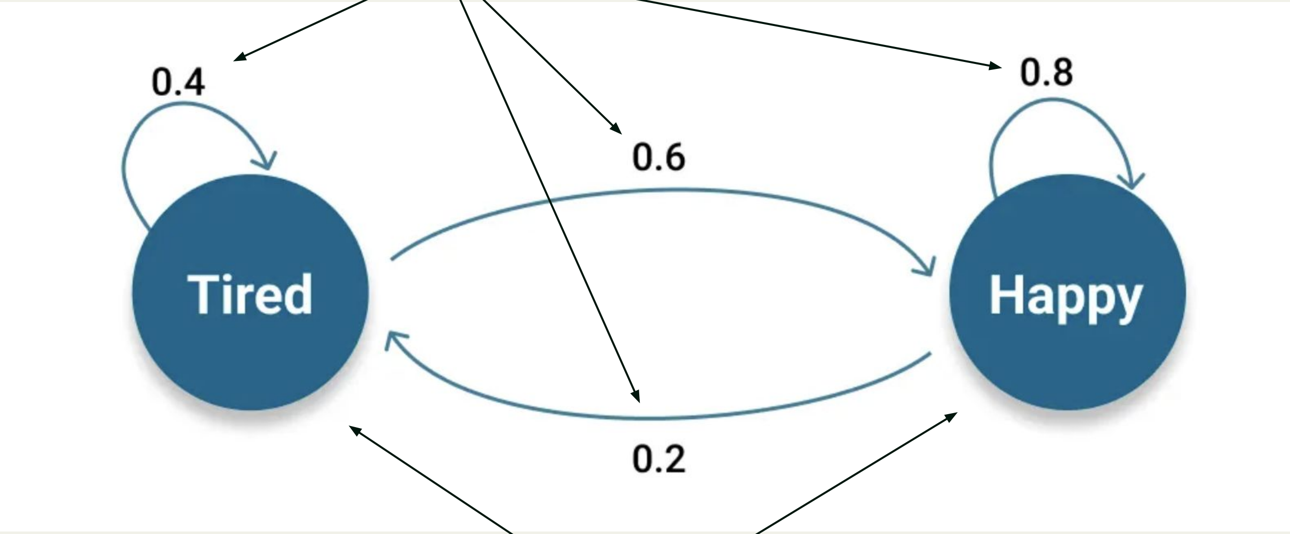


States: the environment/system

- stochastic process
- moves between states over time
- “Memoryless,” The next state only depends on the current state

Markov Chain

Transition Probabilities: from a current state to a next state

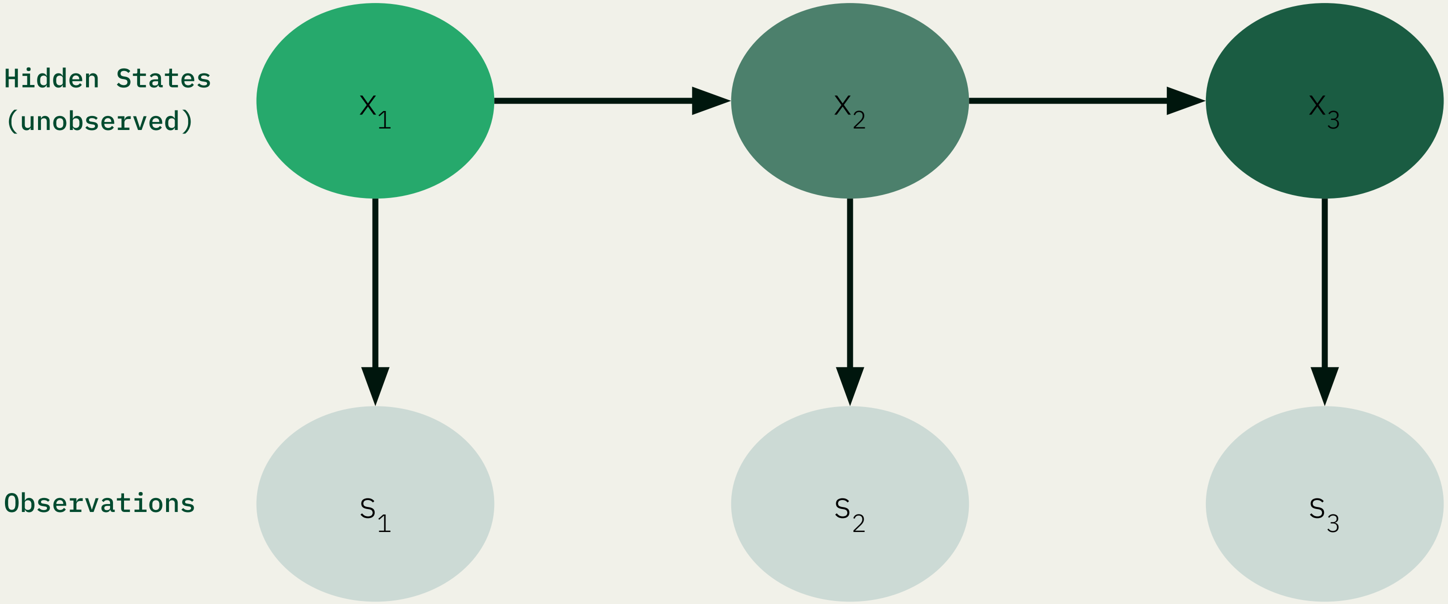


States: the environment/system

Imagine being **Tired** you then have a **60% chance** of becoming happy, and a **40% chance** of staying tired.

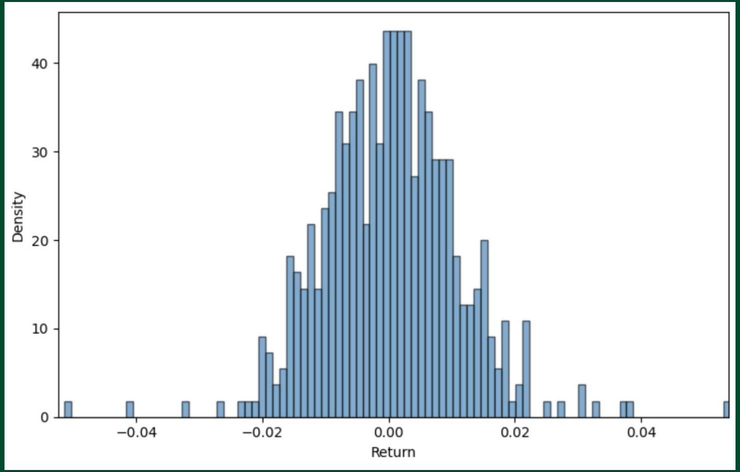
Hidden Markov Model

The Claim: There exists an underlying, unobserved process x_t that governs how s_t are generated.



Stock Returns

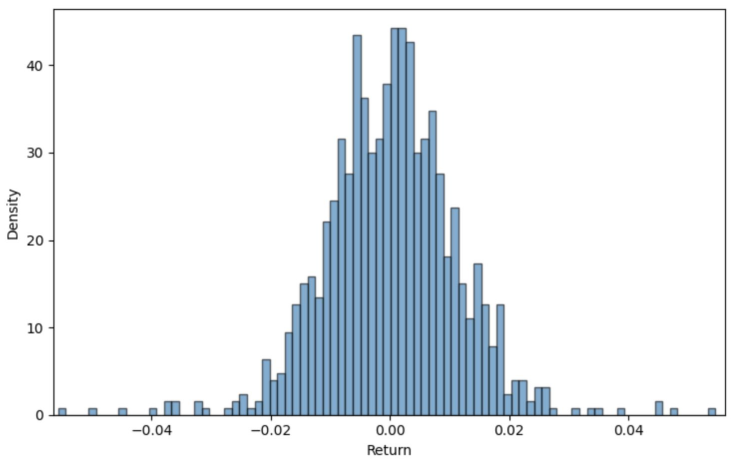
s_t: observed return at time t
n = 500



Goal: Find the true distribution generating these observations

Stock Returns

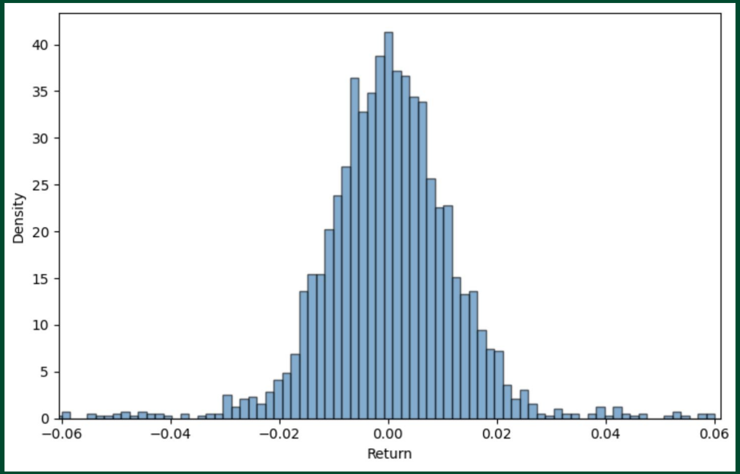
n = 1000



Goal: Find the true distribution generating these observations

Stock Returns

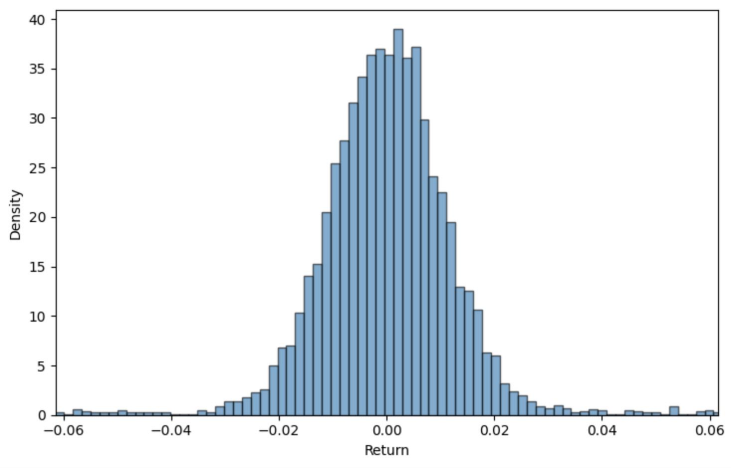
n = 2500



Goal: Find the true distribution generating these observations

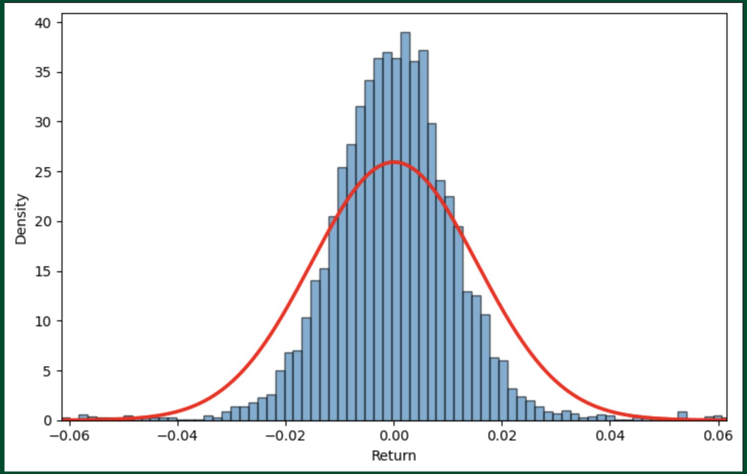
Stock Returns

n = 5000



Goal: Find the true distribution generating these observations

Distribution Generating Returns



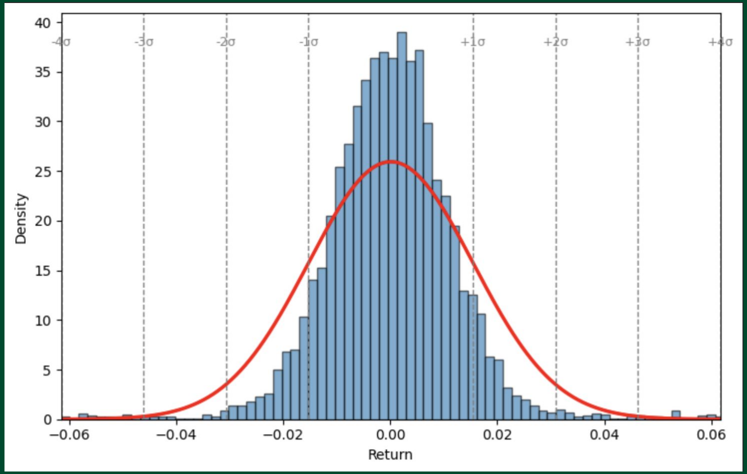
This distribution determines:

- how spread out returns are (volatility)
- how often large moves occur

Relationship: $S \sim N(u, o^2)$

Each day's return is a random draw from a Normal(Gaussian) distribution

68-95-99.7 Rule



In a normal distribution:

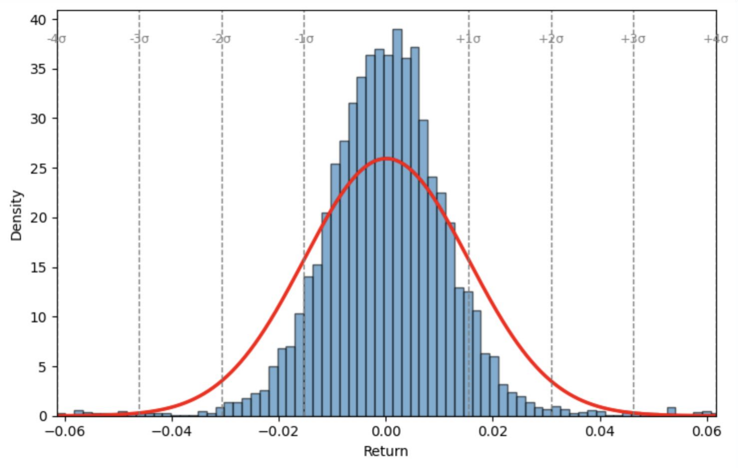
~68% of observations fall within ±1σ

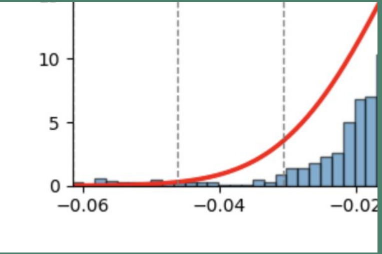
95% fall within ±2σ

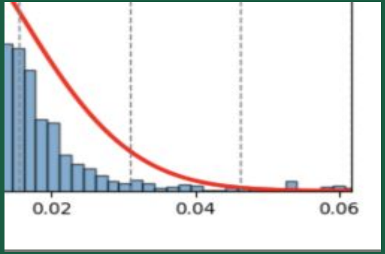
99.7% fall within ±3σ

Relationship: $S \sim N(u, o^2)$

What’s the big deal?

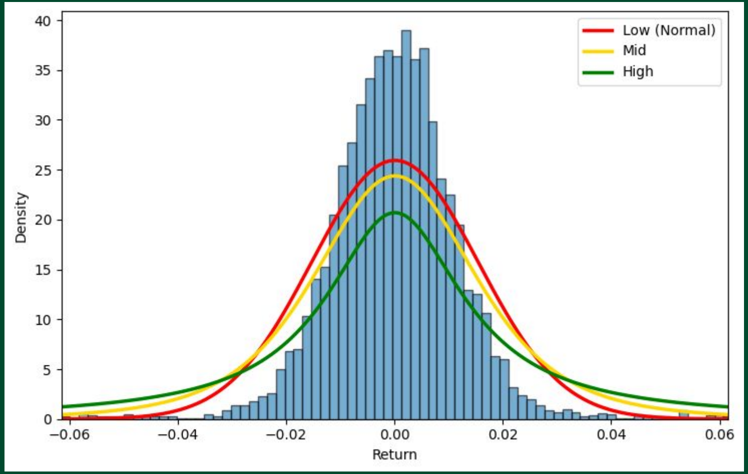






Extreme returns appear too frequently to be explained by a single Gaussian distribution.

Volatility Regimes



Observed Data: Returns

Hidden States: Regimes

Model: Hidden Markov Models

Relationship: $S \sim N(u, o_t^2)$

u: fixed | **o**: varies (like sliding window)

Can a 3-state Hidden Markov Model fitted to daily equity returns identify underlying volatility regimes?

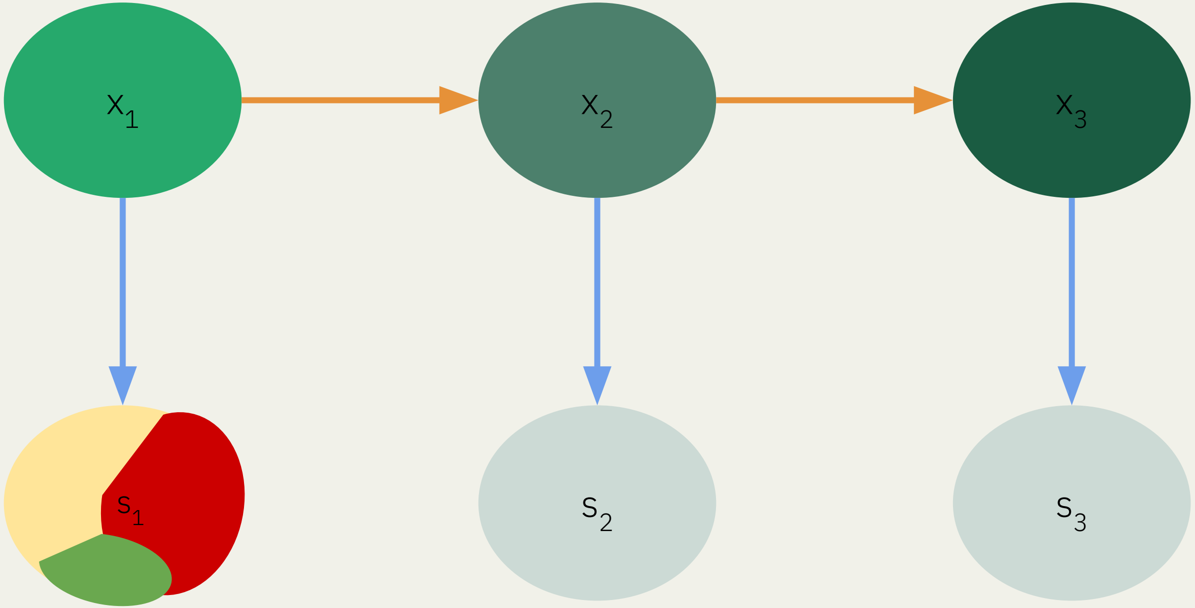
Do those regimes have meaningfully different risk characteristics and "realistic" transition dynamics?

What I've Done So Far

Transition Probability: Probability the Regime change from x_t to x_{t+1}

Emission Probability: Distribution of Returns generated by regime x_t

Possible Regimes (x_t)



Aside: Likelihood of this return under each Regime

Anticipated Challenges / Potential Future Steps

Challenges:

01 Choose number of regimes

02 Sensitivity to initializations

03 Interpreting regimes economically

Next Steps: Examine, for each day, which hidden state the model believes the market was most likely in **(Viterbi Algorithm)**

COMPs Timeline

Now - Feb 16th

Experiment with number of regimes (K = 2, 3, 4)

Evaluate Model strength (precision/recall)

Feb 17th - Mar 19th

Refine HMM implementation

Create more HMM models using other volatility techniques or distributions

Mar 20th - Mar 25th

Results & model comparison results

Start writing up a report on findings

Mar 26th - Apr 7th

Improve visualizations and storytelling

Interpret economic meaning of regimes

Apr 6th- Apr 28th

Is this Viable…
Finalized analysis & conclusions

Cleaning up Github repo(s)

Questions?

Cited Sources

Machine Learning: A Probabilistic Perspective (Kevin P. Murphy), pp. 589–618

https://probml.github.io/pml-book/book1.html

“Hidden Markov Models Explained with a Real-Life Example and Python Code”

https://medium.com/data-science/hidden-markov-models-explained-with-a-real-life-example-and-python-code-2df2a7956d65

YouTube lecture: Markov Models / HMM intuition

https://www.youtube.com/watch?v=9yl4XGp5OEg&list=PLFHg16PBga2kbr7QPGrwsvSnq676RIMBa&index=10

Jurafsky & Martin, Speech and Language Processing

https://web.stanford.edu/~jurafsky/slp3/A.pdf

YouTube lecture: HMM likelihoods and math walkthrough

https://www.youtube.com/watch?v=s9dU3sFeE40