Kenyon College - Feb 3rd, 2026

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: Applied Mathematics concentration in Computer Science

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

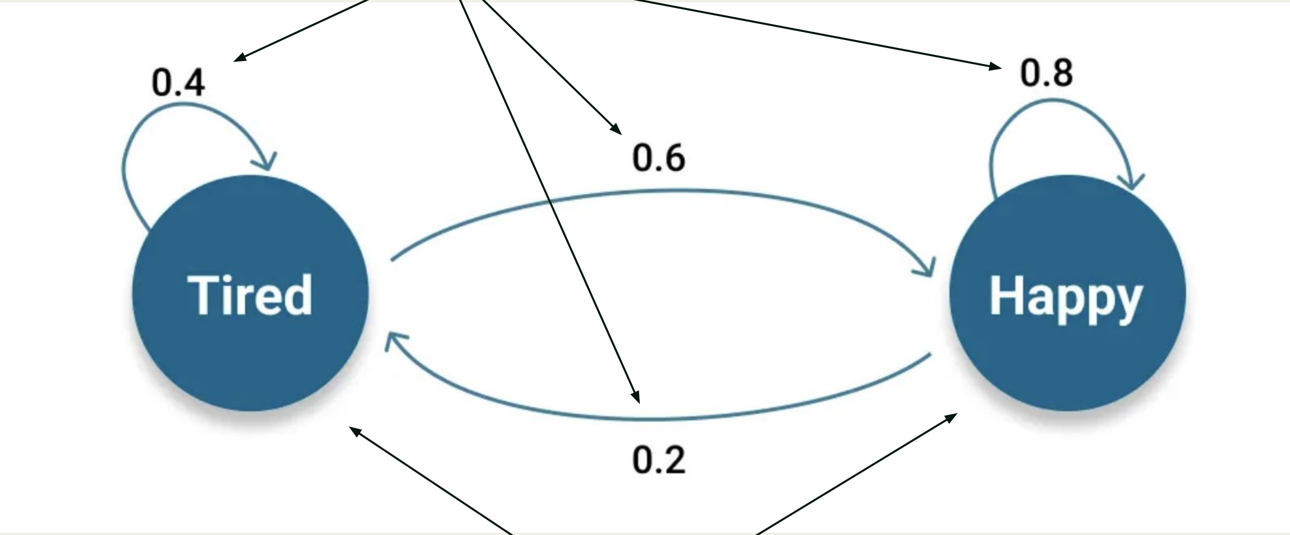
From Data Science → Quant Research

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 Chains

Transition Probabilities: from a current state to a next state



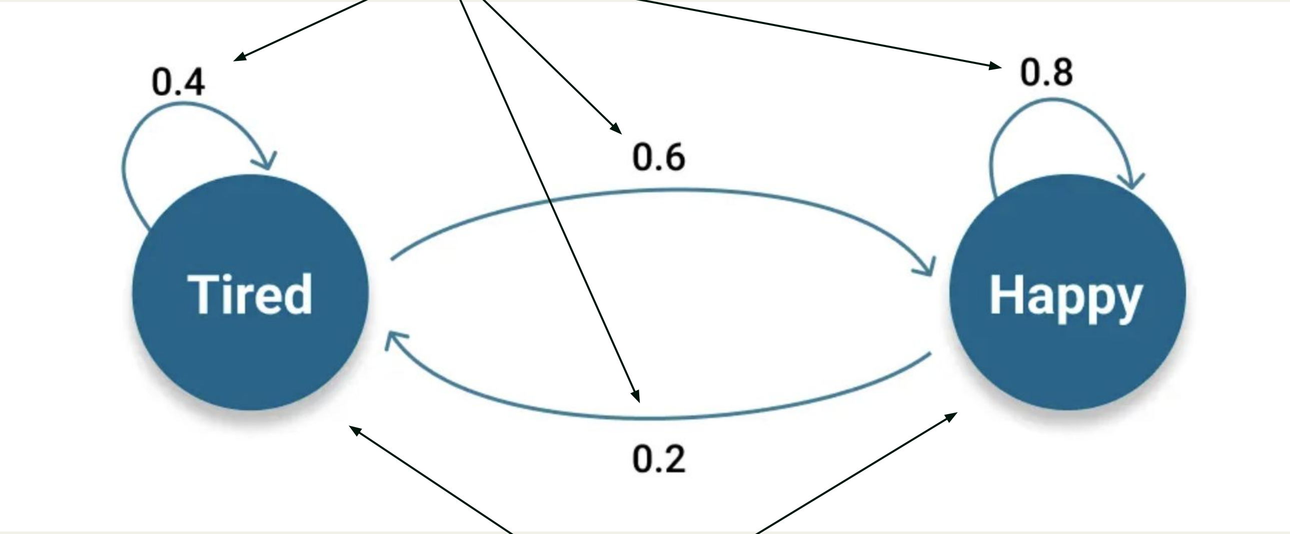
States: the environment/system

Example:

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

Markov Chains

Transition Probabilities: from a current state to a next state

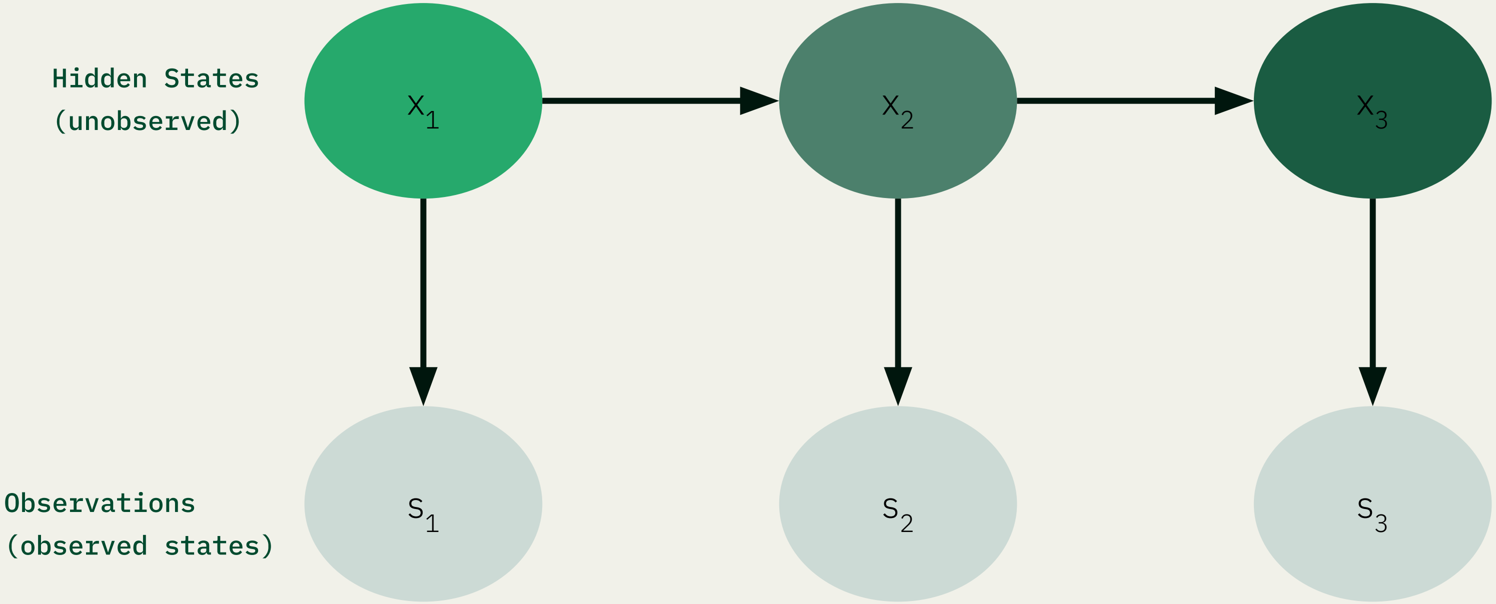


States: the environment/system

- stochastic process
- moves between states over time
- **“Memoryless,”** The next state only depends on the current state. This is what’s called the **Markov Property**

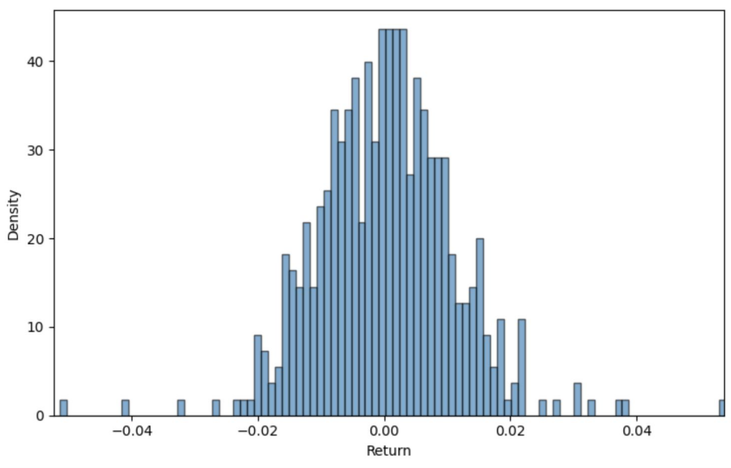
Hidden Markov Model

The Claim: There exists an underlying, unobserved processes $\mathbf{x}_t$ that governs how $\mathbf{s}_t$ are generated.



Stock Returns

n: the number of samples (n=500)
s_t: observed return at time t



This histogram shows the **empirical distribution of observed returns.**

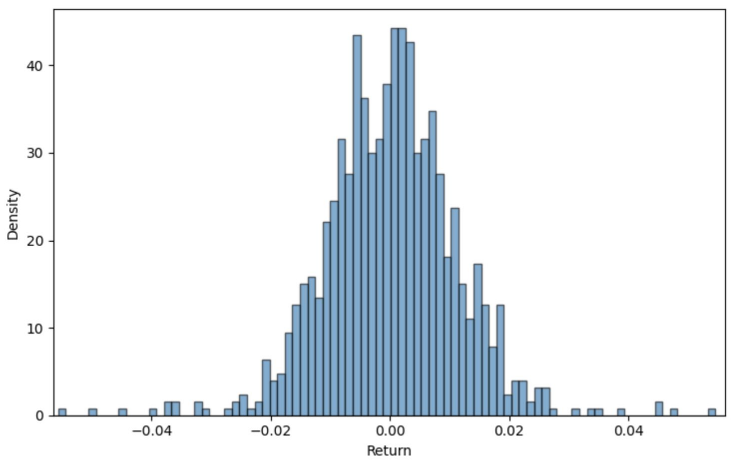
Each **return is a sample** drawn from an **unknown underlying distribution.**

With a **small number of samples**, the shape of the distribution is noisy and difficult to identify, so let's **increase the number of samples.**

Goal: Find the distribution generating these observations **(Find x_t)**

Stock Returns

n = 1000



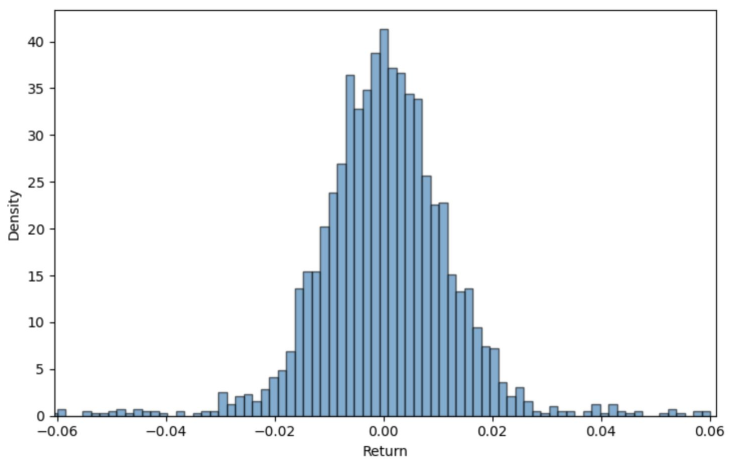
Let's keep increasing the sample count..

We are starting to get **more density around ± [0.02, 0.04]**

Goal: Find the distribution generating these observations **(Find x_t)**

Stock Returns

n = 2500



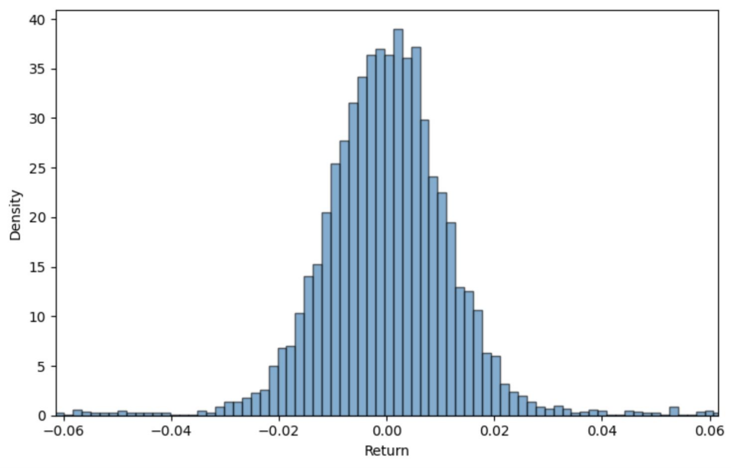
Let's keep increasing the sample count..

Again, we are getting **more density around ± [0.02, 0.04]** and also **less gaps in our structure**

Goal: Find the distribution generating these observations **(Find x_t)**

Stock Returns

n = 5000



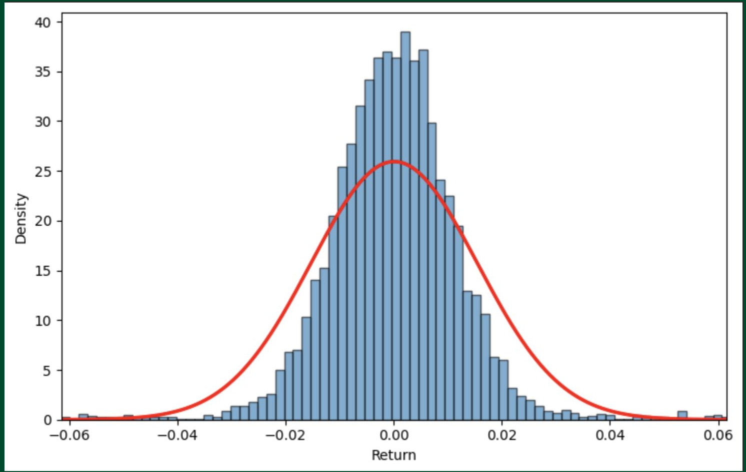
We have 10x our sample size (n)

Our data, now looks **continuous**, meaning **no holes or gaps.**

So, let's try fitting this to a familiar **probability distribution**

Goal: Find the distribution generating these observations **(Find x_t)**

Distribution Generating Returns



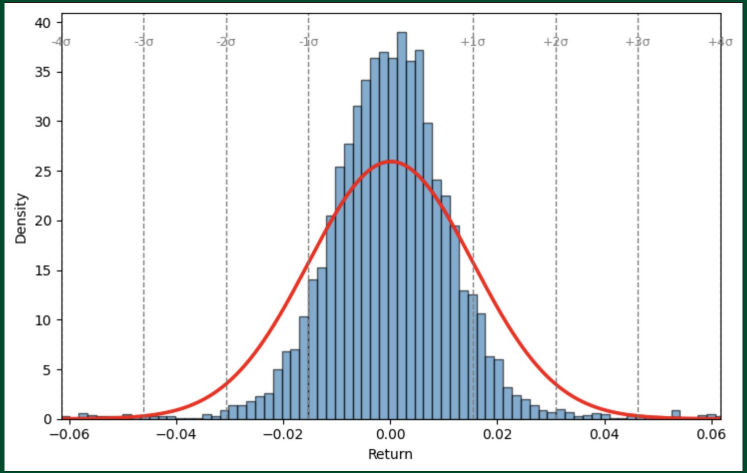
Each day's a return (s_t) is a randomly drawn from a Normal (Gaussian) distribution (x_t)

This distribution determines:

- The **probability** of observing different returns
- The **spread** of returns (volatility)
- The **likelihood** of extreme market moves

Relationship: $S \sim N(\mu, \sigma^2)$
μ : mean (the center of the curve)
σ : standard deviation (volatility)

68-95-99.7 Rule



In a Normal distribution:

~68% of observations fall within ±1σ

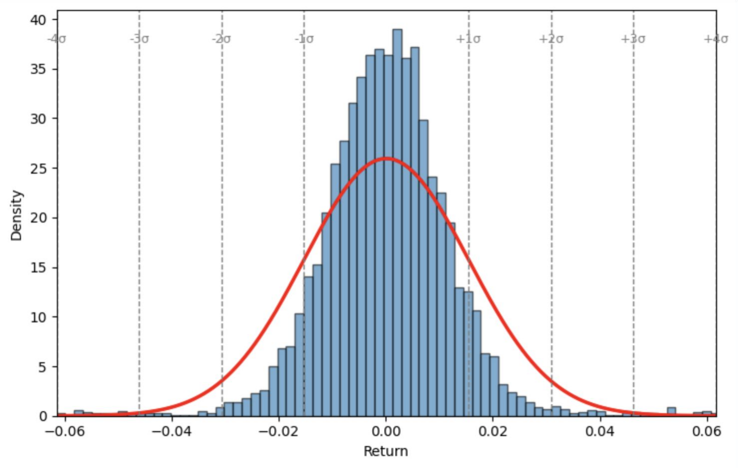
95% fall within ±2σ

99.7% fall within ±3σ

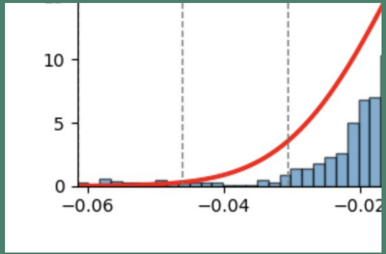
This **Empirical Rule** saves us a lot of time, and tells us that this is **not satisfying our goal**

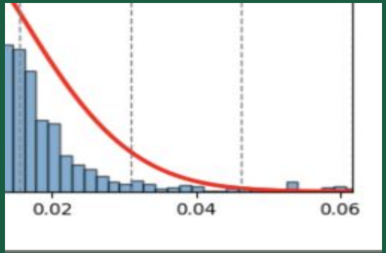
Relationship: $S \sim N(\mu, \sigma^2)$
μ : mean (the center of the curve)
σ : standard deviation (volatility)

What’s the big deal?



Yes, we may be misfitting the center of the distribution. However, to a **portfolio manager**, the **main concern is risk** not the **average return**





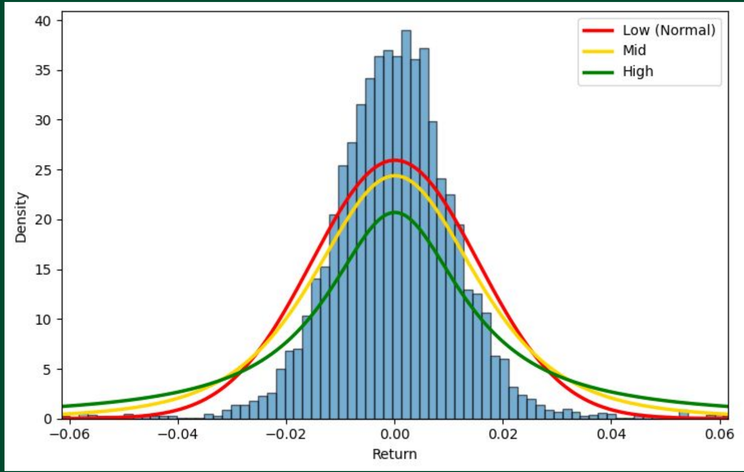
<u>What’s happening at the tails:</u>

Extreme returns occur far more **frequently** than a single **Gaussian distribution predicts**

Volatility Regimes

New Claim:

Returns are generated from **different Gaussian** distributions depending on the underlying **hidden volatility regime**



Old Claim $S \sim N(\mu,\sigma^2)$ → New Claim $S \sim N(\mu,\sigma_{x\_t}^{2})$

Observed Data: Returns ($\mathbf{s_t}$)

Hidden States: Regimes ($\mathbf{x_t}$)

Model: Hidden Markov Model

Parameters with New Claim:

$\boldsymbol{\mu}$: (fixed), $\boldsymbol{\sigma_{x\_t}}$: (regime-dependent **volatility**)

"

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

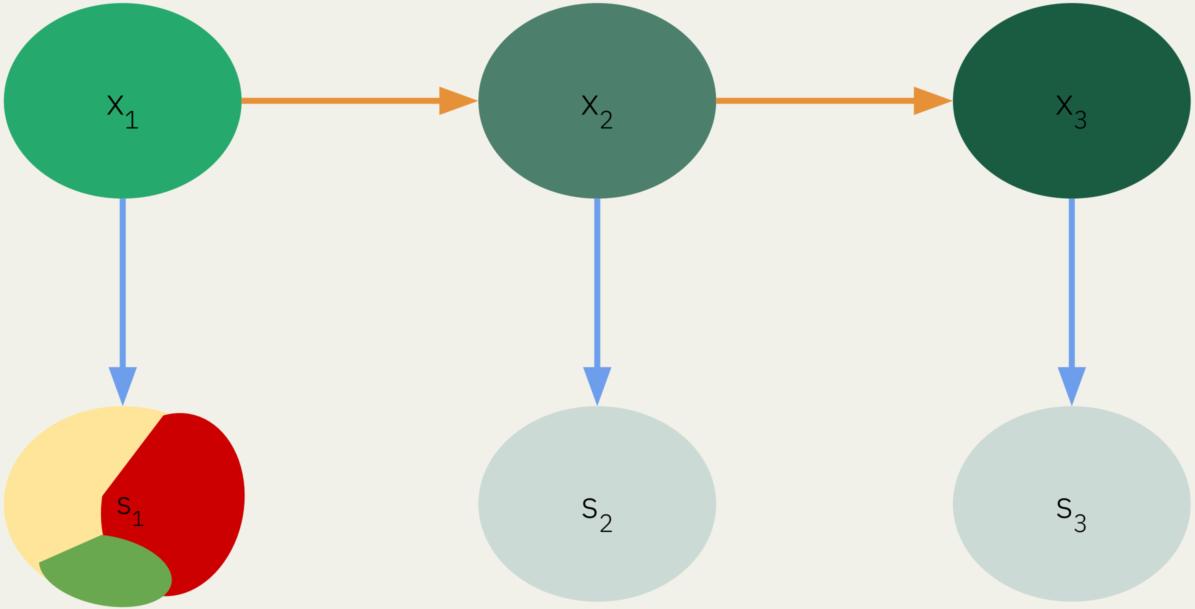
Do those regimes have meaningfully different risk characteristics and "realistic" transition dynamics?

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	Feb 17th - Mar 19th	Mar 20th - Mar 25th	Mar 26th - Apr 7th	Apr 6th- Apr 28th
Experiment with number of regimes (K = 2, 3, 4) Evaluate Model strength (precision/recall)	Refine HMM implementation Create more HMM models using other volatility techniques or distributions	Results & model comparison results Start writing up a report on findings	Improve visualizations and storytelling Interpret economic meaning of regimes	Is this Viable… Finalized analysis & conclusions Cleaning up Github repo(s)

Questions?