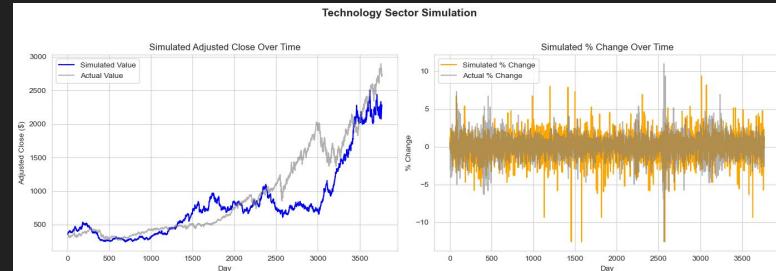
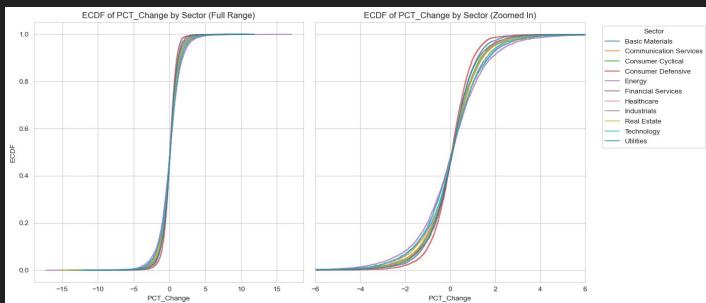


# Stock Price Analysis by Sector

Joseph Hudson, Joseph Kaminetz, Will Novak, Will Brannock, Aiden Rocha, Garret Knapp, Sam Kunitz-Levy



# We created stock projections by sector with Kaggle data

## Data

- Kaggle data sourced from FRED & Yahoo Finance, 2010-2024
- Only used the 150 companies with fully complete data
- All 11 sectors are represented with at least 5 companies after the filtering process

## Analysis

- Conducted analyses by-sector
- 1st technique: Markov chain
- 2nd technique: Randomly sample percent change using ECDF

Stocks data has some missing data but there were enough companies in each sector for a Markov chain

### 1. Missing data is consistent by year

Year	Date	Symbol	Adj Close	Close	High	Low	Open	Volume	Year
2010	0	0	88574	88574	88574	88574	88574	88574	0
2011	0	0	87993	87993	87993	87993	87993	87993	0
2012	0	0	86656	86656	86656	86656	86656	86656	0
2013	0	0	86733	86733	86733	86733	86733	86733	0
2014	0	0	85932	85932	85932	85932	85932	85932	0
2015	0	0	85436	85436	85436	85436	85436	85436	0
2016	0	0	85081	85081	85081	85081	85081	85081	0
2017	0	0	84195	84195	84195	84195	84195	84195	0
2018	0	0	84085	84085	84085	84085	84085	84085	0
2019	0	0	83848	83848	83848	83848	83848	83848	0
2020	0	0	83996	83996	83996	83996	83996	83996	0
2021	0	0	83664	83664	83664	83664	83664	83664	0
2022	0	0	83332	83332	83332	83332	83332	83332	0
2023	0	0	82939	82939	82939	82939	82939	82939	0
2024	0	0	81241	81241	81241	81241	81241	81241	0



### 2. Mix of partially and fully missing data

Symbol	Date	Symbol	Adj Close	Close	High	Low	Open	Volume	Year
A	0	0	3768	3768	3768	3768	3768	3768	0
AAPL	0	0	3768	3768	3768	3768	3768	3768	0
ABBV	0	0	754	754	754	754	754	754	0
ABNB	0	0	3768	3768	3768	3768	3768	3768	0
ABT	0	0	0	0	0	0	0	0	0
...	...	...	...	...	...	...	...	...	...
XYL	0	0	449	449	449	449	449	449	0
YUM	0	0	3768	3768	3768	3768	3768	3768	0
ZBH	0	0	3768	3768	3768	3768	3768	3768	0
ZBRA	0	0	3768	3768	3768	3768	3768	3768	0
ZTS	0	0	3768	3768	3768	3768	3768	3768	0



### 3. Average missingness\* is ~65% across sectors

Sector	missing	Symbol	missing_percentage
Basic Materials	15	22	68.181818
Communication Services	10	22	45.454545
Consumer Cyclical	37	55	67.272727
Consumer Defensive	25	37	67.567568
Energy	17	22	77.272727
Financial Services	42	67	62.686567
Healthcare	41	62	66.129032
Industrials	43	70	61.428571
Real Estate	18	31	58.064516
Technology	59	82	71.951220
Utilities	23	32	71.875000



### Sector

Basic Materials	7	Healthcare	17
Communication Services	9	Industrials	22
Consumer Cyclical	16	Real Estate	12
Consumer Defensive	12	Technology	17
Energy	5	Utilities	9
Financial Services	24	Name: Symbol, dtype: int64	

**Takeaway:** We decided there are enough companies with fully complete data in each sector for by-sector analysis (150 total across 11 sectors)

\* Missingness includes companies with partially and fully missing data

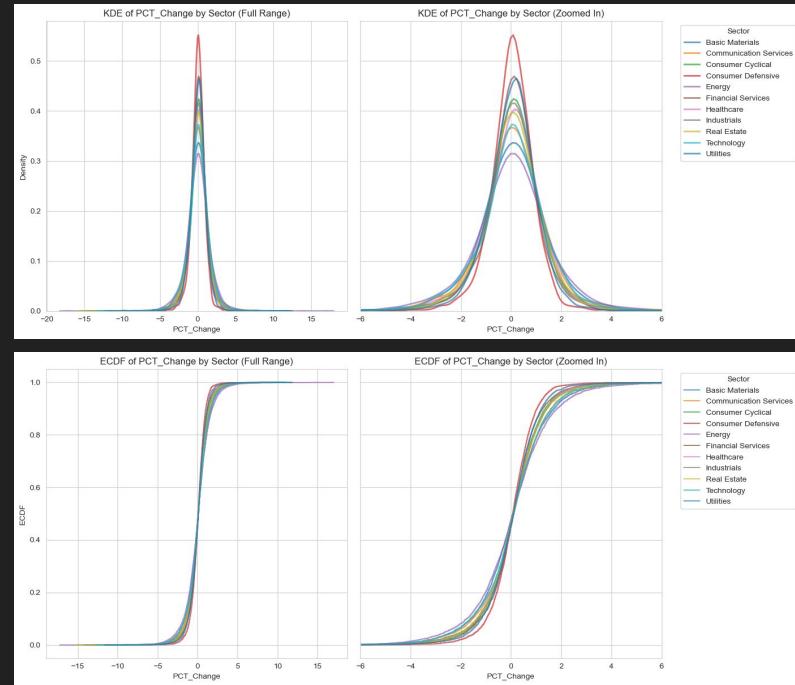
# We calculated daily change and flagged an “up indicator”

Table 1: Adjusted close by sector from 2010-2024

	Date	Sector	Adj Close
0	2010-01-04	Basic Materials	178.189838
1	2010-01-04	Communication Services	128.564411
2	2010-01-04	Consumer Cyclical	241.193282
3	2010-01-04	Consumer Defensive	256.358244
4	2010-01-04	Energy	104.954098
...	...	...	...
41443	2024-12-20	Healthcare	4061.319965
41444	2024-12-20	Industrials	5892.510040
41445	2024-12-20	Real Estate	2011.869986
41446	2024-12-20	Technology	2744.200005
41447	2024-12-20	Utilities	682.539993

Table 2: Percent change with “up indicator”

	Date	Sector	Adj Close	pctChange	up_ind
0	2010-01-04	Basic Materials	178.189838	NaN	0
1	2010-01-04	Communication Services	128.564411	NaN	0
2	2010-01-04	Consumer Cyclical	241.193282	NaN	0
3	2010-01-04	Consumer Defensive	256.358244	NaN	0
4	2010-01-04	Energy	104.954098	NaN	0
...	...	...	...	...	...
41443	2024-12-20	Healthcare	4061.319965	0.015142	1
41444	2024-12-20	Industrials	5892.510040	0.014285	1
41445	2024-12-20	Real Estate	2011.869986	0.012776	1
41446	2024-12-20	Technology	2744.200005	0.013798	1
41447	2024-12-20	Utilities	682.539993	0.015080	1



Sectors central tendency is slightly above 0. Most percent change values are in [-4%, 4%] range

# Model 1: Short-term and long-term Markov chains

## Key details

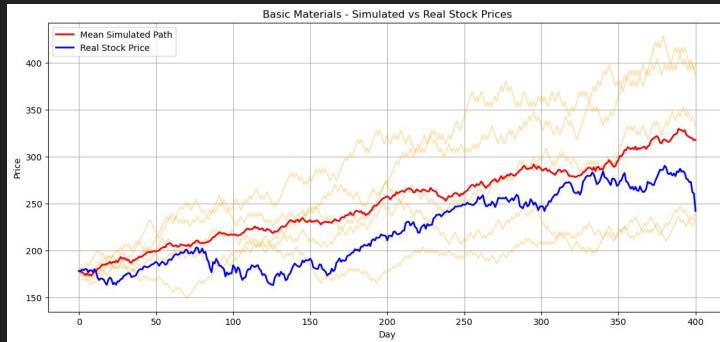
- Short term (400 days)
- Long-term (3,000 days)
- Up penalty:  $e^{\sigma_{\%Change}} (\sim 1.011)$
- Down penalty:  $e^{-\sigma_{\%Change}} (\sim 0.989)$

## ? Key questions

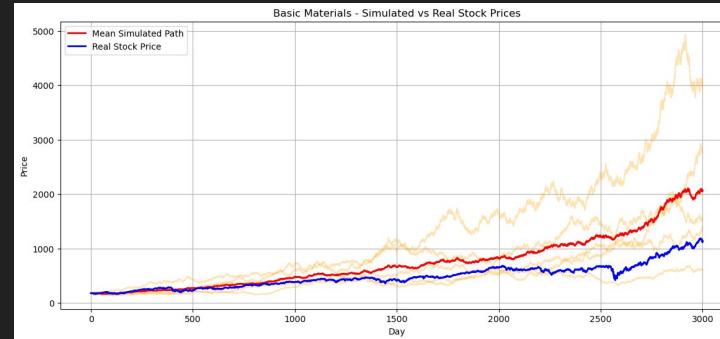
- How do we determine whether the stock should go up or down?
- How much does the stock price go up or down each day?
- Why does geometric decay matter?

# Model 1: Ex. Sectors, Basic Materials and Real Estate

*Markov chain: Short-term (400 days)*



*Markov chain: Long-term (3,000 days)*



# Model 2: Random sampling from percent change distribution

## Method

---

1. Start with each sector's adjusted close on the first day
2. Select a percent change from the distribution
3. Use the percent change to recalculate next day's adj. close
4. Repeat until horizon (short and long term) is met



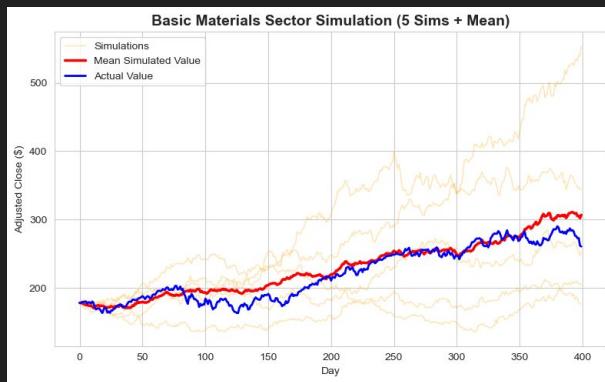
## Challenges

---

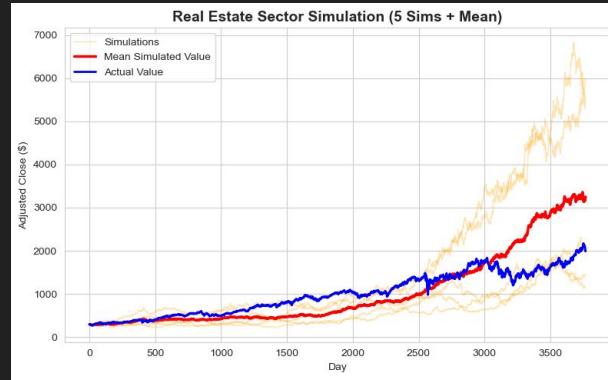
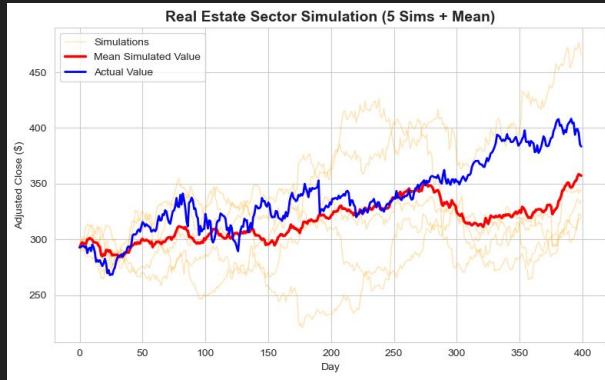
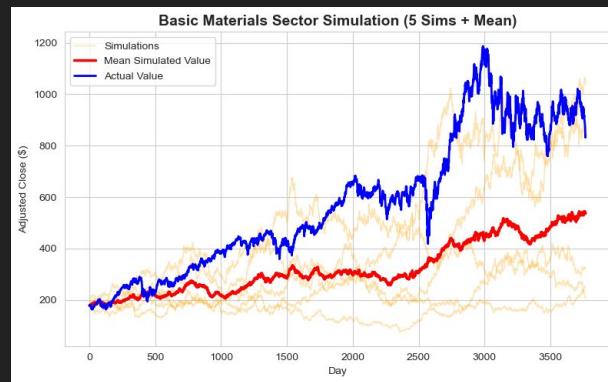
- How many paths should we simulate?
- How to aggregate the simulated paths?

# Model 2: Random Sampling Approach

*Short-term (400 days)*

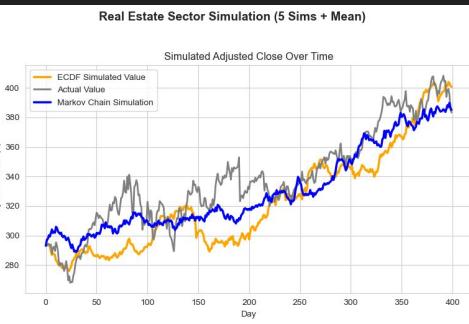


*Long-term (3,000 days)*



# Cross model evaluation

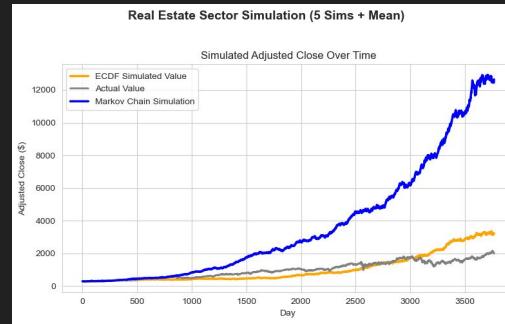
Real estate



Basic Materials



## Long-term (3,769 days)



## Takeaways

- Performance differs by sector
- Both models perform better in the short term
- Markov chain goes to steady state in long term
- Sampling struggles to capture the fluctuations in price



# Conclusions and Limitations

## Conclusions

- Hard to exactly mirror a sectors closing value, but can achieve reasonable ballpark
- ECDF performs better long-term, but the overall method fails to consider ordering of percent change
- It is possible to model short term momentum changes, but hard to mirror the magnitude of change



## Limitations

- Missing stock data (65%)
- We only evaluated Model performance visually
- Markov chain does not vary the magnitude of percent change
- ECDF-based model has tendency to underfit data as simulations increase
- Neither model fully accounts for future volatility or uncertainty

# Potential Future Work

## Future Work

---

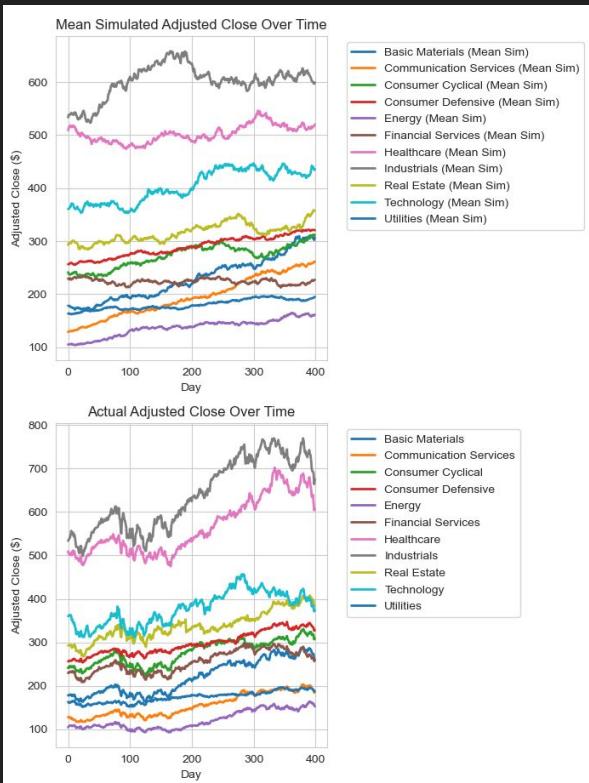
- Changing how we calculated up and down factor for our Markov chains
- Changing how the sample for percent change from ECDF model is picked
- Comparing our ECDF and Markov Chain models with other financial models like time Series

# Appendix Slides

The End!

# Model 2 Predicted vs Actual

## Short Term



## Long Term

