Here is the introduction current introduction to the article. It's an excellent start but we might need to change it given I’ve finalized the asset classes to include. I’d also like you to expand the “current introduction” to include **“New Content to thematically integrate into introduction” –** not word for word but thematically. I’ve also left several prompts in place where I need your help; they look like this “<CHATGPT….” – Fill this in please. So make this all into a new introduction and then with this and your knowledge of the code so far, suggest the structure of how I should finish the article.

* **“Current Introduction”**

Title: "Quantifying Market Efficiency: A Python & Machine Learning Approach Across Asset Classes":

In the vast expanse of the financial universe, every asset class has its own distinct rhythm. Like the unique patterns in a heartbeat, each tells its own story, mirroring the complex web of economic activities, investor sentiments, geopolitical influences, and myriad factors that drive global finance. But how do these rhythms relate? Do they dance in sync or chart their own courses? In this article, we'll embark on a journey across different asset classes, gauging their efficiency, their synchronicity, and ultimately, their potential opportunities. From the frenetic pace of equities to the sturdier march of bonds, and from the volatility of commodities to the subtleties of currencies, let's unravel the tales these asset classes tell.

* **“New Content to thematically integrate into introduction:”**

Working at a predominantly fixed income asset manager most of my career, I would consistently hear the portfolio managers and research analyst discuss “trading relative value” up and down a firm’s capital structure to capture alpha or excess returns – from senior secured assets like high yields bonds to the equity. This suggests that there are market inefficiencies between different securities in a given firm’s capital structure – but I had never seen quantitative support of this claim. This led me to wonder if this was true in a broader sense, i.e. are there differences in market efficiency across major asset classes? Can we use machine learning to answer this at the asset class level, then at an individual security level and finally devise a simplistic trading strategy to exploit these inefficiencies.

While there are decades of peer reviewed research by professionals far more experienced than I am, I wanted to utilize the predictive analytics tools at my disposal and industry expertise to attempt to address these questions. There will be many assumptions and holes in this analysis and while feedback is welcome, this work does not claim to be the end-all on this topic.

First, I will describe the different layers of market efficiency and why they’re important in the context of active management. Then using various python libraries (including blp for the Bloomberg Professional Service), we empirically quantify which asset classes potential exhibit inefficiencies during the analysis period, examine a few single company examples and then explore potential ML driven investment strategies.

All view are my own and past performance is not indicative of future results <CHATGPT finish please>……

* Add brief bullets describing in simple terms the Weak form, Semi-strong and Strong form efficient market hypothesis in the context of active management.

**Throw in some references to Burton Malkiel’s random walk work**

**Metrics to include in introduction discussion :**

* Weak-form: auto-correlation
* Semi-Strong form: <ChatGPT insert metrics>
* Strong form: <ChatGPT insert metrics>

- These are the current indexes for analysis and their names:

tickers = ['RUITR Index','RU20INTR Index', 'C0A0 Index','H0A0 Index','SPBDAL Index', 'MXEA Index', 'MXEF Index','EMUSTRUU Index']

readable\_names = ['US Large Cap Equities','US Small Cap Equities','US Invest.Grade Bonds', 'US High Yield Bonds', 'US Bank Loans', 'Developed Country Equities', 'Emerging Market Equities','Emerging Market Debt']

This is very good. Revised the below pargraphs to make them less corny (less casual) and slightly more professional and slightly shorter.

## Should I also include “The Landscape of Asset Returns”,” Unraveling Drawdown Dynamics: An Efficiency Paradox?” and a correlation analysis? If so, where should they go in the structure?

”

Revise:

My career in a predominantly fixed income asset management firm has been an eye-opener. Conversations around the water cooler often revolved around trading relative value within a firm's capital structure to seize alpha. It was a dance of navigating from high yield bonds at the top, right down to the equity, suggesting potential inefficiencies along the way. But where was the quantitative proof? This lingering question morphed into a broader quest: Do market efficiencies truly differ across major asset classes? And if so, could machine learning help us unearth, quantify, and then exploit these inefficiencies?

While the corridors of academia are lined with theses on this subject, my aim is to combine the analytical prowess of machine learning with my industry insights to unravel this enigma. There will undoubtedly be chinks in the armor - assumptions, possible oversights, but this is a starting point, an exploratory endeavor rather than a conclusive statement. Feedback, as always, is a gift.

### The Landscape of Asset Returns

In assessing the return distributions over approximately 16 years of daily returns, certain asset classes showcase notable characteristics hinting at inefficiencies. For instance, the exceptionally high kurtosis of **Oil** suggests it experiences extreme return movements, while **US Bank Loans** have both a significant negative skew and a heightened kurtosis, implying potential left-tailed events and extreme fluctuations. **US High Yield Bonds** also exhibit a pronounced negative skew, hinting at potential downside risks. On the other hand, the mild skewness and kurtosis of assets like **US Aggregate Bonds and European Equities** might suggest they behave more in line with a normal distribution. Such distributions provide a basis to further explore underlying factors or anomalies, setting the stage for our deeper dive into measures of efficiency.

## Unraveling Drawdown Dynamics: An Efficiency Paradox?

Drawing insights from the largest drawdowns of various asset classes offers a nuanced perspective on market efficiency. US High Yield Bonds and US Bank Loans, for instance, display noticeable signs of potential inefficiency in their return distributions. Intriguingly, they also register the smallest drawdowns, which might seem counterintuitive at first. Smaller drawdowns often indicate a more stable asset class, but when combined with other inefficiencies, they raise questions. Could inefficiencies, in some cases, play a stabilizing role, or are there other underlying factors at play? These findings certainly set the stage for more detailed investigations.

From an investment strategy standpoint, the stability of these assets, combined with their potential inefficiencies, makes them attractive candidates. It hints at possible opportunities to capitalize on mispricings while also enjoying a relatively stable performance. The following sections will delve deeper into designing investment strategies around these insights.

### Correlation between Asset Classes:

US Large Cap Equities US Small Cap Equities \

US Large Cap Equities 1.000000 0.917423

US Small Cap Equities 0.917423 1.000000

US Invest.Grade Bonds -0.142154 -0.157417

US High Yield Bonds 0.416106 0.372620

US Bank Loans 0.245718 0.211915

Developed Country Equities 0.529898 0.474994

Emerging Market Equities 0.474016 0.424242

Emerging Market Debt 0.242080 0.216006

US Invest.Grade Bonds US High Yield Bonds \

US Large Cap Equities -0.142154 0.416106

US Small Cap Equities -0.157417 0.372620

US Invest.Grade Bonds 1.000000 0.317015

US High Yield Bonds 0.317015 1.000000

US Bank Loans 0.265066 0.729963

Developed Country Equities 0.029995 0.604484

Emerging Market Equities 0.002984 0.555244

Emerging Market Debt 0.508022 0.639150

US Bank Loans Developed Country Equities \

US Large Cap Equities 0.245718 0.529898

US Small Cap Equities 0.211915 0.474994

US Invest.Grade Bonds 0.265066 0.029995

US High Yield Bonds 0.729963 0.604484

US Bank Loans 1.000000 0.363448

Developed Country Equities 0.363448 1.000000

Emerging Market Equities 0.366793 0.782296

Emerging Market Debt 0.504951 0.435250

Emerging Market Equities Emerging Market Debt

US Large Cap Equities 0.474016 0.242080

US Small Cap Equities 0.424242 0.216006

US Invest.Grade Bonds 0.002984 0.508022

US High Yield Bonds 0.555244 0.639150

US Bank Loans 0.366793 0.504951

Developed Country Equities 0.782296 0.435250

Emerging Market Equities 1.000000 0.460264

Emerging Market Debt 0.460264 1.000000

Can you rewrite the below paragraph given the updated table below and the suggestion you gave for this section in the outline?

Outline suggestion from you:

The Landscape of Asset Returns: Before diving into the efficiency analysis, take a step back to lay out the landscape of asset returns. Here, you can offer a historical perspective, trend analysis, and set the baseline for subsequent discussions.

Original: The Landscape of Asset Returns

In assessing the return distributions over approximately 16 years of daily returns, certain asset classes showcase notable characteristics hinting at inefficiencies. For instance, the exceptionally high kurtosis of \*\*Oil\*\* suggests it experiences extreme return movements, while \*\*US Bank Loans\*\* have both a significant negative skew and a heightened kurtosis, implying potential left-tailed events and extreme fluctuations. \*\*US High Yield Bonds\*\* also exhibit a pronounced negative skew, hinting at potential downside risks. On the other hand, the mild skewness and kurtosis of assets like \*\*US Aggregate Bonds and European Equities\*\* might suggest they behave more in line with a normal distribution. Such distributions provide a basis to further explore underlying factors or anomalies, setting the stage for our deeper dive into measures of efficiency.

New data:

Skewness Kurtosis

US Large Cap Equities -0.291443 11.402500

US Small Cap Equities -0.374045 6.802141

US Investment Grade Bonds -0.718321 7.101279

US High Yield Bonds -1.589386 26.072258

US Bank Loans -4.352170 93.225062

Developed Country Equities -0.359710 8.526835

Emerging Market Equities -0.313688 8.225888

Emerging Market Debt -2.004527 27.623686

Can you draft up a summary paragraph for the “Correlation Analysis” portion of the article? I’ve included the suggested structure and content you gave earlier?

Data:

US Large Cap Equities US Small Cap Equities \

US Large Cap Equities 1.000000 0.917423

US Small Cap Equities 0.917423 1.000000

US Invest.Grade Bonds -0.142154 -0.157417

US High Yield Bonds 0.416106 0.372620

US Bank Loans 0.245718 0.211915

Developed Country Equities 0.529898 0.474994

Emerging Market Equities 0.474016 0.424242

Emerging Market Debt 0.242080 0.216006

US Invest.Grade Bonds US High Yield Bonds \

US Large Cap Equities -0.142154 0.416106

US Small Cap Equities -0.157417 0.372620

US Invest.Grade Bonds 1.000000 0.317015

US High Yield Bonds 0.317015 1.000000

US Bank Loans 0.265066 0.729963

Developed Country Equities 0.029995 0.604484

Emerging Market Equities 0.002984 0.555244

Emerging Market Debt 0.508022 0.639150

US Bank Loans Developed Country Equities \

US Large Cap Equities 0.245718 0.529898

US Small Cap Equities 0.211915 0.474994

US Invest.Grade Bonds 0.265066 0.029995

US High Yield Bonds 0.729963 0.604484

US Bank Loans 1.000000 0.363448

Developed Country Equities 0.363448 1.000000

Emerging Market Equities 0.366793 0.782296

Emerging Market Debt 0.504951 0.435250

Emerging Market Equities Emerging Market Debt

US Large Cap Equities 0.474016 0.242080

US Small Cap Equities 0.424242 0.216006

US Invest.Grade Bonds 0.002984 0.508022

US High Yield Bonds 0.555244 0.639150

US Bank Loans 0.366793 0.504951

Developed Country Equities 0.782296 0.435250

Emerging Market Equities 1.000000 0.460264

Outline suggestion from you:

After setting the stage with historical returns, it's a natural transition to explore how these asset classes correlate with each other. This section can be pivotal in understanding co-movement patterns and potential diversification benefits.

Next up is “Empirical Analysis Across Asset Classes: Weak Form Efficiency and Autocorrelation.” I need you to review the below output from the autocorrelation charts and provide some insight in paragraph form similar to the correlation analysis above. I’ve provided a prior output you wrote on an old copy of the article. Make it one or two paragraphs, citing key metrics and such.

Autocorrelations:

US Large Cap Equities ACF: [ 1.00000000e+00 -1.21112288e-01 4.07607563e-03 1.10647571e-02

-4.38805302e-02 -6.48053071e-03 -4.14112507e-02 4.73496924e-02

-4.01193225e-02 5.50758494e-02 -5.64270413e-04]

------------------------------------

US Small Cap Equities ACF: [ 1. -0.10780908 0.04037373 -0.01429901 -0.03536153 -0.00855236

-0.03562108 0.04140333 -0.02463358 0.03083223 0.01828271]

------------------------------------

US Invest.Grade Bonds ACF: [1.00000000e+00 7.85037491e-02 3.60584222e-02 5.69982731e-02

1.79399274e-02 1.54051001e-02 1.89856368e-02 4.66807012e-04

1.39102390e-02 6.37902302e-04 4.30887733e-02]

------------------------------------

US High Yield Bonds ACF: [1. 0.41725202 0.266062 0.19299907 0.1427607 0.12227732

0.03427907 0.08119397 0.08082896 0.10573247 0.09276863]

------------------------------------

US Bank Loans ACF: [1. 0.55883271 0.45934017 0.42900577 0.28020686 0.26301483

0.12437095 0.20641112 0.15768224 0.10376255 0.0888124 ]

------------------------------------

Developed Country Equities ACF: [ 1. 0.10826241 -0.02450978 -0.02814654 0.01996633 -0.03211051

-0.03170518 0.02631902 0.0069439 -0.02023705 -0.01522399]

------------------------------------

Emerging Market Equities ACF: [ 1. 0.16598092 0.03284168 -0.01133032 -0.02528721 -0.02628881

-0.04757302 0.04591439 -0.00701511 0.00569924 -0.00665568]

------------------------------------

Emerging Market Debt ACF: [1. 0.48381411 0.31659704 0.21591137 0.12979926 0.07019519

0.00126961 0.02672043 0.04234336 0.02953238 0.05970234]

------------------------------------

### Prior content generated for article on same section: Insights by Asset Class

* **US Aggregate Bonds**: Displayed very little autocorrelation, suggesting a high level of efficiency in the pricing of this asset class.
* **US High Yield Bonds**: Displayed a significant positive autocorrelation at lag 1, suggesting potential momentum in returns. This might hint at inefficiencies or behavioral biases among traders.
* **US Equities**: Showed negative autocorrelation, potentially hinting at mean reversion in this market.
* **US Leveraged Loans**: Strong positive autocorrelations for several lags suggest some predictability in returns, challenging the notion of efficient markets.
* European Equities & Asian Equities: These markets showed very little autocorrelation, suggesting that they might be close to a random walk.
* **Oil**: The positive autocorrelation at lag 1 suggests momentum, while the negative values at later lags hint at potential mean reversion.
* Gold: Displayed a random walk characteristic with most autocorrelations close to zero.
* EUR/USD: Shows signs of a random walk with minor inefficiencies.
* **US Real Estate**: The negative autocorrelation at lag 1 may hint at some mean reversion in this market.

Actually, take the below code and add the following ticker/name. then save a separate csv file for later in the same folder called “sentiment\_scores.csv” – there should be a date column and then score column in this csv. After that that file is saved, delete that column from the original ‘index\_data\_raw’ dataframe and execute the rest of the file as planned. There should be no sentiment\_score in the index\_data\_raw’ dataframe.

Code to add:

tickers= 'SFFRNEWS Index'

readable\_name='Sentiment Score'

Code to change:

from xbbg import blp

import pandas as pd

import numpy as np

import seaborn as sns

import matplotlib

import matplotlib.pyplot as plt

from scipy.stats import skew, kurtosis,bartlett

from statsmodels.tsa.stattools import acf

import statsmodels.api as sm

from datetime import datetime, timedelta

import os

import warnings

warnings.simplefilter(action='ignore', category=FutureWarning)

# Set the font properties

plt.rcParams['font.family'] = 'sans-serif'

plt.rcParams['font.sans-serif'] = ['Arial']

%matplotlib inline

# Log into Bloomberg Terminal before executing

tickers = ['RUITR Index','RU20INTR Index', 'C0A0 Index','H0A0 Index','SPBDAL Index', 'MXEA Index', 'MXEF Index','EMUSTRUU Index',]

readable\_names = ['US Large Cap Equities','US Small Cap Equities','US Invest.Grade Bonds', 'US High Yield Bonds', 'US Bank Loans', 'Developed Country Equities', 'Emerging Market Equities','Emerging Market Debt',]

csv\_file\_path = "data/index\_data.csv"

# Check if CSV file exists

if not os.path.exists(csv\_file\_path):

def fetch\_data(tickers, start\_date, end\_date):

return blp.bdh(tickers=tickers, flds=['Px\_Last'], start\_date=start\_date, end\_date=end\_date)

# start\_date = datetime.today() - timedelta(days=365\*23) # Last 20 years data

# end\_date = datetime.today()

date\_string = "4/2/2007"

date\_format = "%m/%d/%Y"

start\_date = datetime.strptime(date\_string, date\_format)

end\_date = datetime.today()

index\_data\_raw = fetch\_data(tickers, start\_date, end\_date)

# Rename columns for better readability

index\_data\_raw.columns = readable\_names

threshold = len(index\_data\_raw.columns) - 2

# Drop rows with more than threshold missing values

index\_data = index\_data\_raw.dropna(thresh=threshold)

# Save data to CSV

index\_data.to\_csv(csv\_file\_path)

else:

# Load from CSV if the file already exists

index\_data = pd.read\_csv(csv\_file\_path, index\_col=0, parse\_dates=True)

# Calculate returns

index\_returns = index\_data.pct\_change().dropna()