GROUP WORK PROJECT # 1 MScFE 692: Capstone Review
GROUP NUMBER: **10750**

FULL LEGAL NAME	LOCATION (COUNTRY)	EMAIL ADDRESS	MARK X FOR ANY NON-CONTRIBUTING MEMBER
Gerardo Jose Suarez Rodriguez	Poland	gerardofingurbe@gmail.com	
Darsh A Patel			x
Giorgio Greto	United Kingdom	giorgiogreto@yahoo.com	

Statement of integrity: By typing the names of all group members in the text boxes below, you confirm that the assignment submitted is original work produced by the group (excluding any non-contributing members identified with an “X” above).	
Team member 1	**Gerardo Jose Suarez Rodriguez**
Team member 2	
Team member 3	**Giorgio Greto**

Use the box below to explain any attempts to reach out to a non-contributing member. Type (N/A) if all members contributed. **Note:** You may be required to provide proof of your outreach to non-contributing members upon request.
We created a discord group and had extensive discussions there we pinged him multiple times but no response from him

Group Work Project 1

0.1 Step 1

Extend the PCA monitoring framework in lesson 4.3 with the goal of obtaining a systemic risk indicator combining signals of correlation breakdown and of widening credit spread The computational framework adopted in lesson § 4.3 made use of nine ETFs of the SPDR series, issued by State Street to target specific industry and service sectors in the economy, with each ETF composed exclusively of stocks that are within the S&P500 index. The computations start from January 2007 and extend until the present date. That is why only the nine ETFs already existing in 2007, out of the full suite of eleven SPDR ETFs, were used in the analysis.

Beyond the SPDR ETFs, extra tickers were tracked to follow the volatility of the markets (^VIX), the 10-year Treasury bonds yields (^TNX), and an ETF from iShares reproducing the yield of investment grade corporate bonds (LQD).

All of the data series above are taken with a daily frequency at markets closing time.

Further, we gather daily yield series from the FRED database, for Baa investment grade $\geq$ 20-year corporate bonds as graded by Moody's (BAA) and again the 10-year Treasury bond yields (DGS10).

To summarise:

Ticker	Description
SPDR ETFs	
`XLB`	S&P500 Materials stocks
`XLE`	S&P500 Energy stocks
`XLF`	S&P500 Financial stocks
`XLI`	S&P500 Industrial stocks
`XLK`	S&P500 Technology stocks
`XLP`	S&P500 Consumer Staples stocks
`XLU`	S&P500 Utilities stocks
`XLV`	S&P500 Healthcare stocks
`XLY`	S&P500 Consumer Discretionary stocks
other market data	
`^VIX`	Market volatility index
`^TNX`	10-year Treasury yields index
`LQD`	Liquid, investment-grade corporate bond yields
FRED data	
`BAA`	Moody's Baa-rated corporate bond yields

Ticker	Description
DGS10	10-year Treasury yields

Below, the shape of the `pandas` DataFrame for the 9 SPDR ETFs downloaded, sporting for dimensions the number of daily observations for the period 2007-2025 (after data cleaning) × the 9 ETFs.

```
(4684, 9)
```

We see that the number of observations for the remaining data series extracted from markets agrees with the previous database at 4684 datapoints:

```
(4684, 3)
```

FRED series `BAA`, although nominally taken with daily frequency as declared on the FRED website, in reality exhibits monthly frequency when extracted from the database.

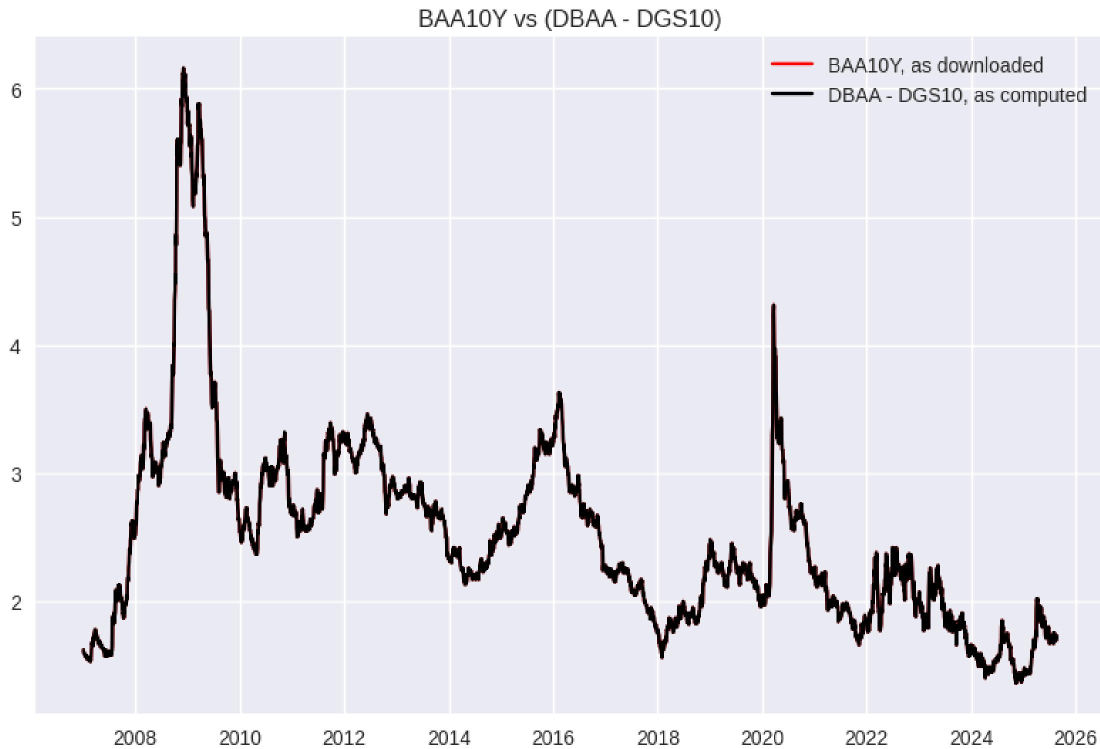
```
id                                                                  BAA
realtime_start                                               2025-08-21
realtime_end                                                 2025-08-21
title                          Moody's Seasoned Baa Corporate Bond Yield
observation_start                                            1919-01-01
observation_end                                              2025-07-01
frequency                                                       Monthly
frequency_short                                                       M
units                                                           Percent
units_short                                                           %
seasonal_adjustment                              Not Seasonally Adjusted
seasonal_adjustment_short                                           NSA
last_updated                                     2025-08-01 10:16:05-05
popularity                                                           71
notes                        These instruments are based on bonds with matu…
dtype: object
```

A workaround is found in downloading FRED's `DBAA` series instead, with the initial letter D evidently standing for *daily*.

Given that the Treasury yields data `DGS10` have more observations than all other series downloaded, in fact spanning over days when markets were closed, we will fill the missing datapoints in `DBAA` with the previous most recent observation available.

There also exist a FRED series `BAA10Y` = `DBAA - DGS10` which tracks the credit spread of investment-grade corporate bonds out of the box. The graph below shows `BAA10Y` is obtained exactly as `DBAA - DGS10`.

```
(4862, 3)
```



For the sake of comparing the credit spread obtained from FRED data (`BAA10Y`) with the one obtained from market data (`LQD - ^TNX`), we first adapt the former to have the same number of observations of the latter.

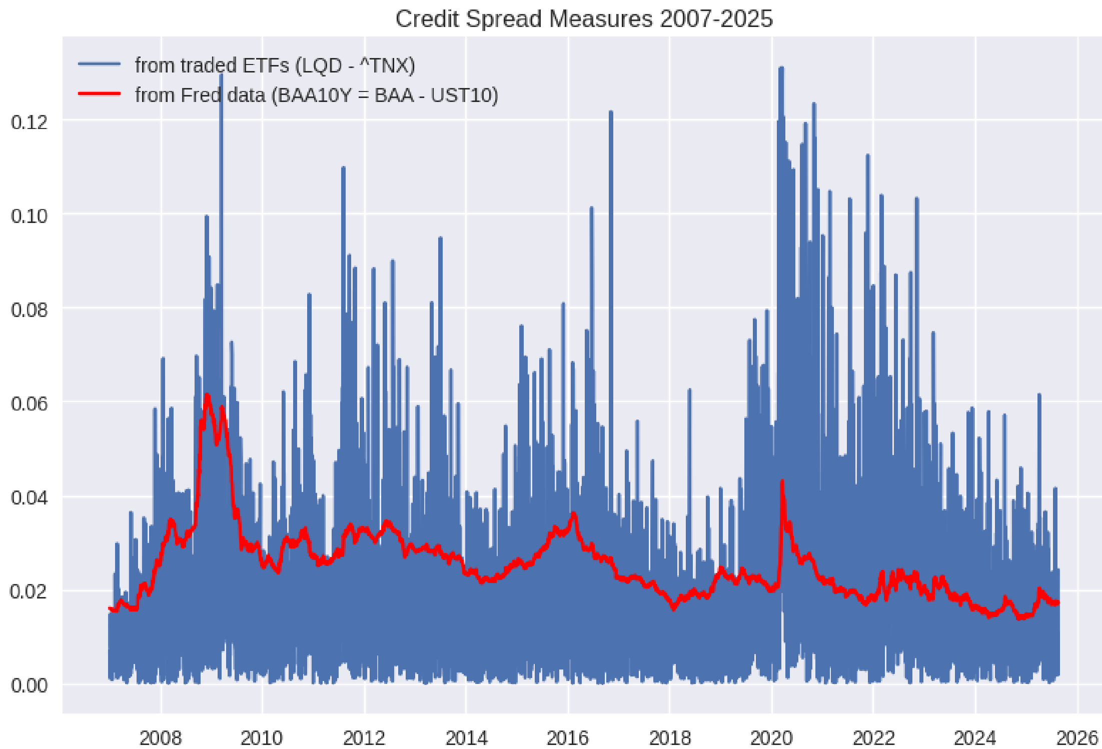
```
shape of spread data from markets:  (4684,)
```

```
shape of spread data from FRED:  (4684,)
```

The credit spreads so obtained by two different sources are plotted in the figure below.

It is evident that the market measure (blue) is much more noisy than that obtained from FRED data (red). We attribute this to the nature of `LQD` as a liquid (highly traded) instrument. Its price is determined by supply and demand dynamics, rather than by the yields of the corporate bonds undergirding it. In turn, this price will reflect the bonds' yields, but only indirectly through market participants' behaviour.

In support of the previous observation, we observe that the absolute value of the noisy market data seems to follow the same patterns as the more regular FRED data. The two measures of credit spread seem to be correlated.



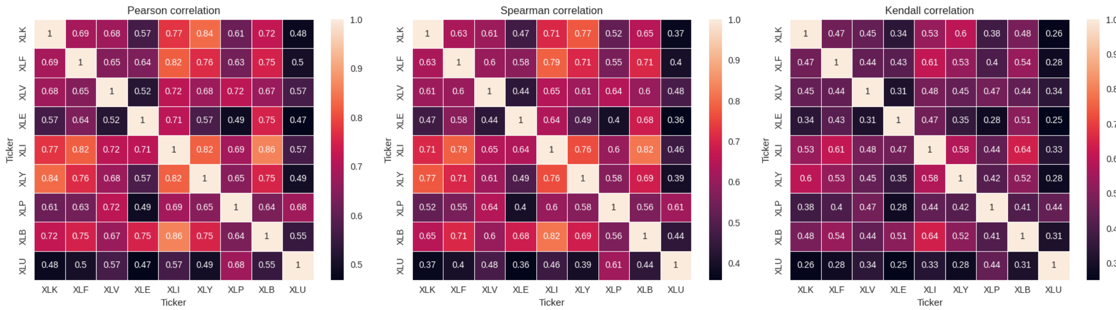
Now, in order to observe the correlations amongst the SPDR ETFs during the selected 2007-25 time window, we compute rolling correlation matrices for time windows of 60 trading days, advancing one day at the time.

The spectral decomposition (or PCA) of these rolling correlation matrices will yield the eigenvalues of the matrices. The largest eigenvalue is found to represent the *market factor* that affects all securities, determining parallel shifts in stock prices.

Its *explained variance ratio* (EVR), i.e. the fraction of total portfolio variance it explains, grows in times of increased market uncertainty. This consists in an undesirable increase in global correlation that could be exploited as a signal for a possibly incipient financial crisis.

We have updated the code in lesson §4.3 to compute the rolling correlation matrices to include not just computing the Pearson (linear) correlation, but also Spearman and Kendall correlations, to track nonlinear effects.

Below, we show the three correlation matrices for the 9 SPDR ETFs, not limited to a few days rolling window, but across the whole time period 2007-25:



We can see from the colorbars in the figures above, showing extreme values, that if nonlinear effects are included (Spearman and Kendall matrices), then the 9 ETFs are overall less correlated than the linear correlation matrix (Pearson) would imply.

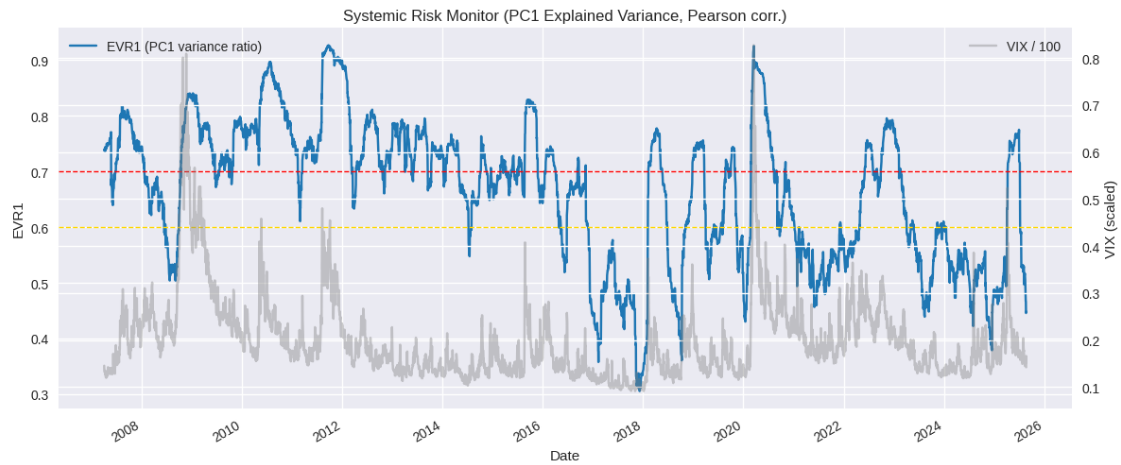
Next, from the rolling 60-days correlations of returns, we compute the EVR for the first eigenvalue (EVR1). We do it for each of the Pearson, Spearman and Kendall correlation matrices.

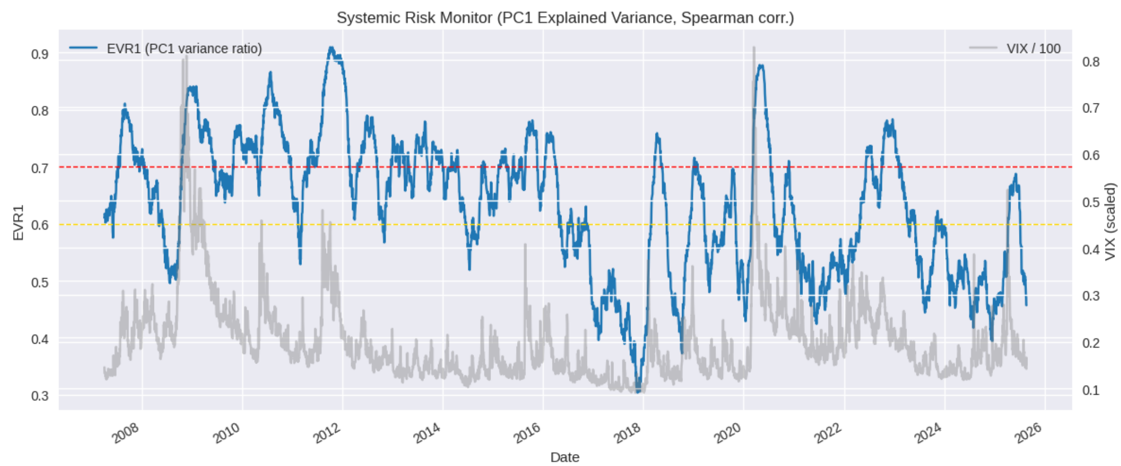
Notice how, on an average home workstation, the computational cost of obtaining the rolling Kendall correlations with Python is 20× more expensive than with the other methods.

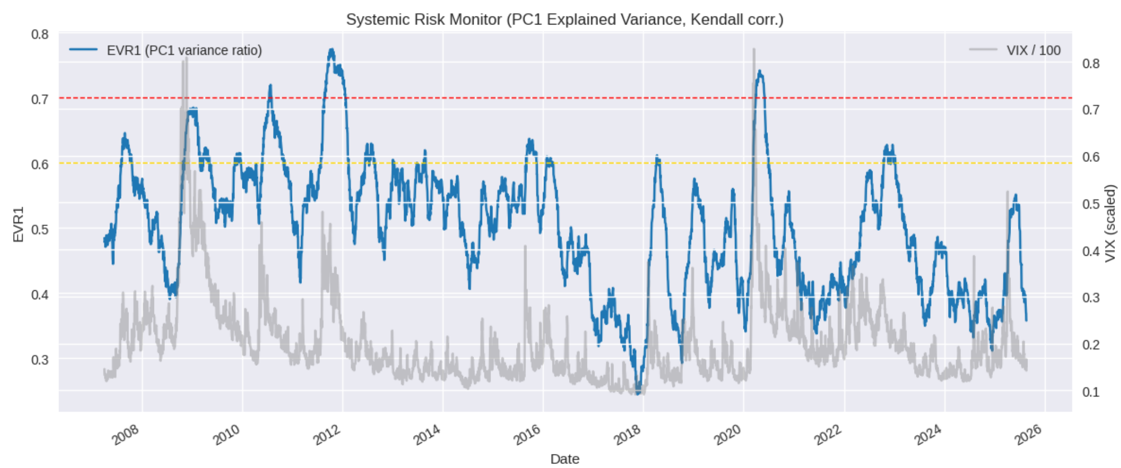
Correlation	Elapsed computation time (s)
Pearson	2.4
Spearman	3.9
Kendall	84.5

From plotting these EVR1s (blue) against time, and in comparison with the ^VIX index (gray), we observe that the EVR1 signal obtained from the Kendall correlation is much more conservative than the previous two. It crosses the highest threshold of correlation (red dotted line) only in three occasions: 2010, 2011-12, and 2020.

It is noticeable that, contrary to the other two correlation methods, the Kendall correlation signal missed the 2007-08 Quant Crash and Great Financial Crisis.







The next step of the analysis consists in checking whether credit spreads and our EVR1 signal are correlated. It would be expected they are as credit spreads widen during systemic crises, just as the proportion of variance explained by the first eigenvalue (EVR1) grows larger.

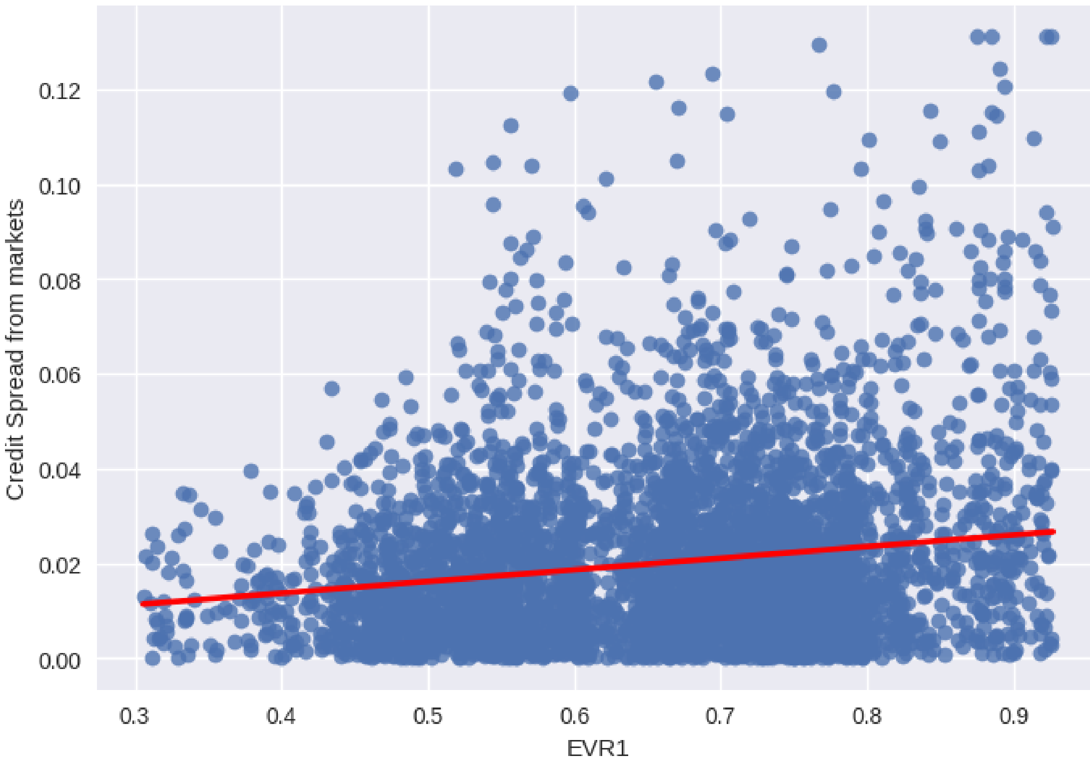
We have two measures of credit spread, one coming from the FRED database and the other downloaded from the markets (LQD ticker as stand-in for corporate bonds), so we have to select the one better attuned to the EVR1 dataset.

For this purpose, we are going to proceed with a linear regression of the Pearson correlation EVR1 data with both the databases of credit spread at our disposal.

```
                                 OLS Regression Results
================================================================================
======
Dep. Variable:     Credit Spread from markets   R-squared:
0.060
Model:                                    OLS   Adj. R-squared:
0.059
Method:                         Least Squares   F-statistic:
98.20
Date:                        Tue, 26 Aug 2025   Prob (F-statistic):
1.28e-61
Time:                                00:46:32   Log-Likelihood:
12062.
No. Observations:                        4625   AIC:
-2.412e+04
Df Residuals:                            4621   BIC:
-2.409e+04
Df Model:                                   3
Covariance Type:                    nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const         -0.0929      0.008    -11.815      0.000      -0.108      -0.077
EVR1           0.1220      0.008     15.141      0.000       0.106       0.138
EVR2           0.1323      0.013     10.017      0.000       0.106       0.158
EVR3           0.2269      0.022     10.371      0.000       0.184       0.270
==============================================================================
Omnibus:                     1728.974   Durbin-Watson:                   1.628
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             7381.371
Skew:                           1.801   Prob(JB):                         0.00
Kurtosis:                       8.033   Cond. No.                         114.
==============================================================================

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

```
<AxesSubplot: xlabel='EVR1', ylabel='Credit Spread from markets'>
```



P-values = 0 for the EVR1-2-3 coefficients in the regression are encouraging, however a low value for the R^2 statistic = 0.060 on the contrary signals that a very low proportion of the total variance of the dependent variable, the credit spread, is explained by the EVR coefficients.

Therefore, we record a mixed outcome for this regression.

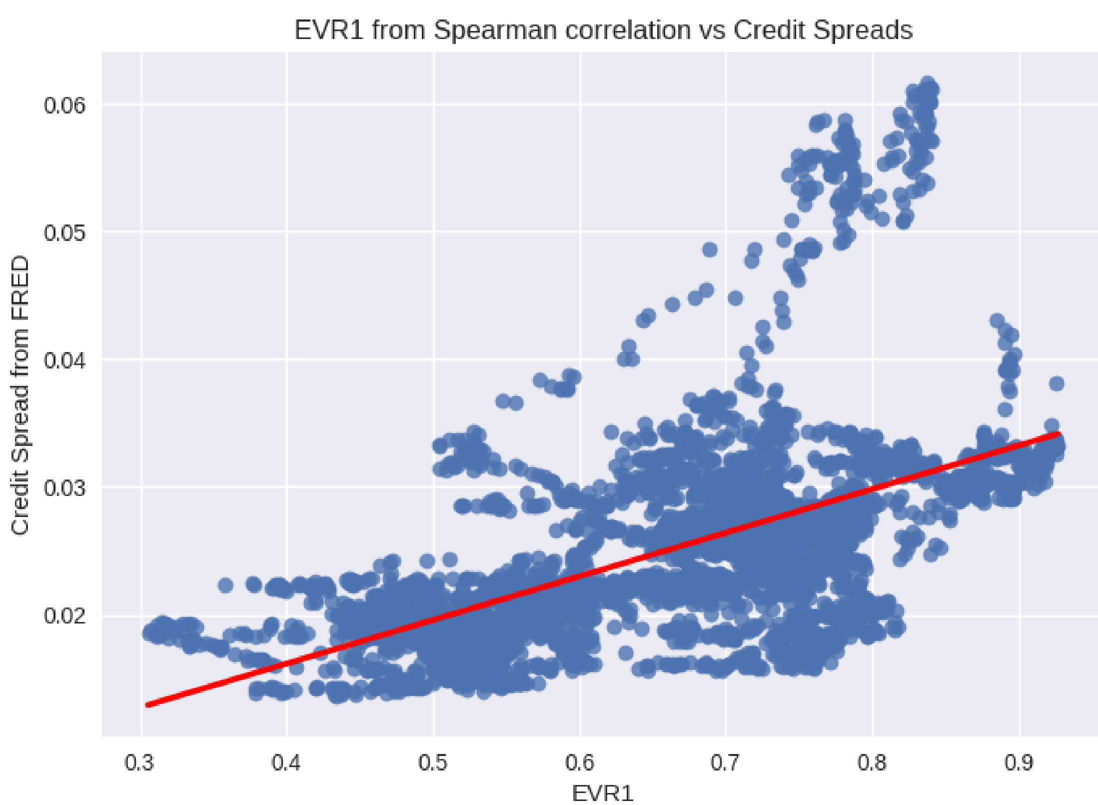
Next, much better results are obtained by regressing the EVR measure against the credit spread series from FRED:

```
                            OLS Regression Results
================================================================================
===
Dep. Variable:     Credit Spread from FRED   R-squared:
0.296
Model:                                 OLS   Adj. R-squared:
0.295
Method:                      Least Squares   F-statistic:
647.1
Date:                     Tue, 26 Aug 2025   Prob (F-statistic):
0.00
Time:                             00:46:33   Log-Likelihood:
16644.
No. Observations:                     4625   AIC:
-3.328e+04
```

```
Df Residuals:                    4621   BIC:
-3.325e+04
Df Model:                           3
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0002      0.003      0.079      0.937      -0.005       0.006
EVR1           0.0364      0.003     12.172      0.000       0.031       0.042
EVR2           0.0093      0.005      1.887      0.059      -0.000       0.019
EVR3          -0.0050      0.008     -0.610      0.542      -0.021       0.011
==============================================================================
Omnibus:                     1886.509   Durbin-Watson:                   0.005
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             9626.096
Skew:                           1.914   Prob(JB):                         0.00
Kurtosis:                       8.942   Cond. No.                         114.
==============================================================================

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

```
Text(0.5, 1.0, 'EVR1 from Spearman correlation vs Credit Spreads')
```



Now the R^2 statistics is more solid, and at the same time the EVR1 coefficient estimation shows a solid p-stat score.

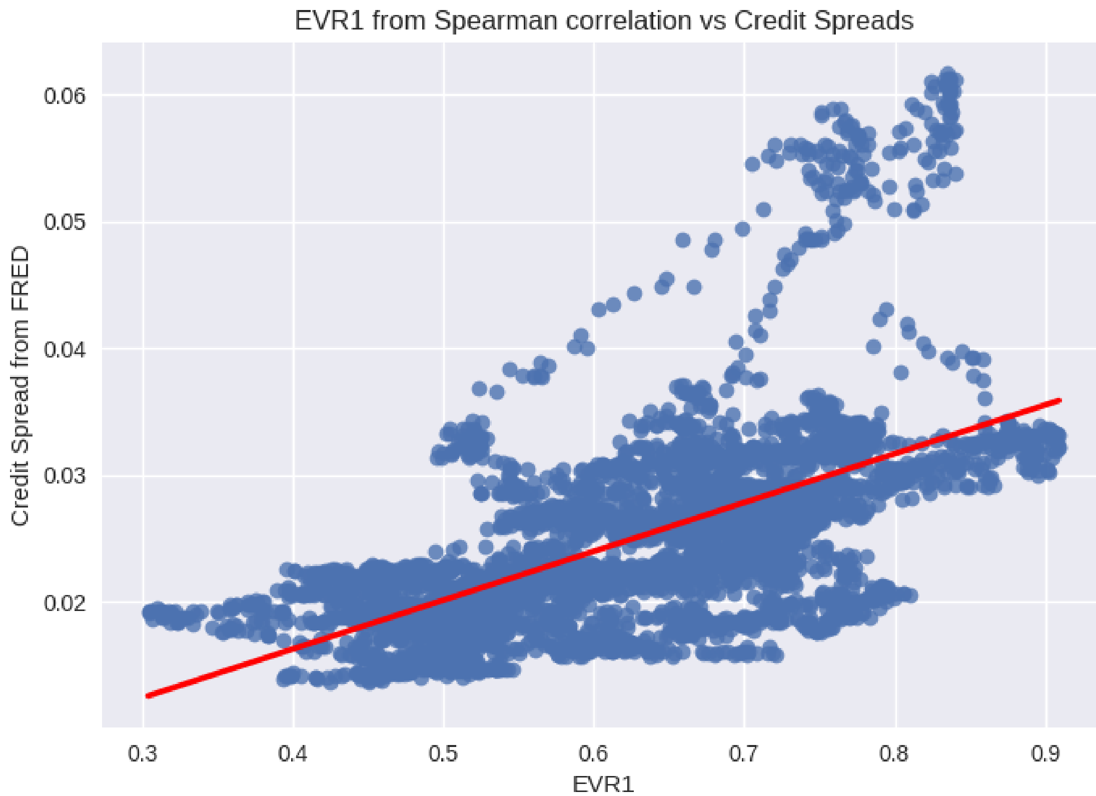
From this information we can infer that credit spreads from Fred are more dependable than their market counterpart, obtained from the LQD ETF.

Proceeding, we complete the picture with the next two regressions which will help selecting the most accurate correlation method for EVR1, in terms of how good it relates to the Fred credit spread.

```
                              OLS Regression Results
================================================================================
===
Dep. Variable:     Credit Spread from FRED   R-squared:
0.346
Model:                                 OLS   Adj. R-squared:
0.346
Method:                      Least Squares   F-statistic:
816.3
Date:                     Tue, 26 Aug 2025   Prob (F-statistic):
0.00
Time:                             00:46:33   Log-Likelihood:
16817.
No. Observations:                     4625   AIC:
-3.363e+04
Df Residuals:                         4621   BIC:
-3.360e+04
Df Model:                                3
Covariance Type:                 nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0003      0.002      0.144      0.885      -0.004       0.005
EVR1           0.0389      0.002     16.164      0.000       0.034       0.044
EVR2           0.0233      0.005      5.079      0.000       0.014       0.032
EVR3          -0.0353      0.007     -4.888      0.000      -0.049      -0.021
==============================================================================
Omnibus:                     1733.807   Durbin-Watson:                   0.005
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             8029.259
Skew:                           1.772   Prob(JB):                         0.00
Kurtosis:                       8.395   Cond. No.                         98.5
==============================================================================

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

```
Text(0.5, 1.0, 'EVR1 from Spearman correlation vs Credit Spreads')
```

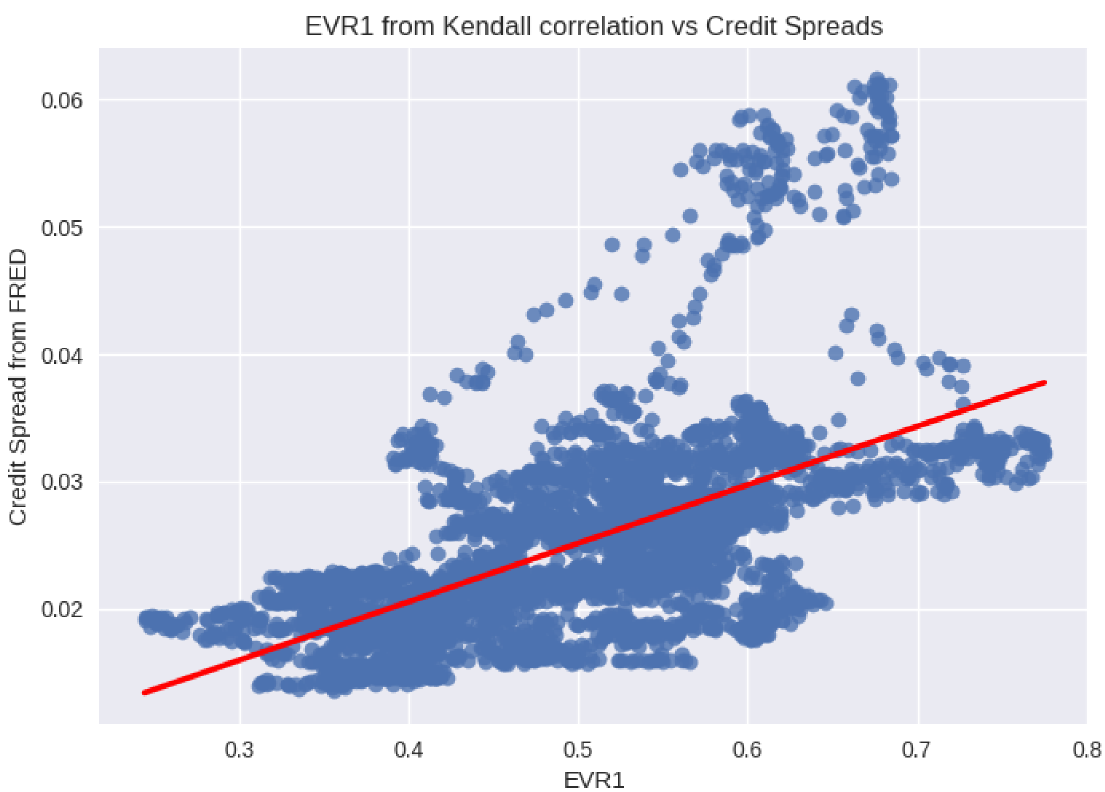



```
                            OLS Regression Results
==============================================================================
===
Dep. Variable:     Credit Spread from FRED   R-squared:
0.349
Model:                                 OLS   Adj. R-squared:
0.348
Method:                      Least Squares   F-statistic:
824.2
Date:                     Tue, 26 Aug 2025   Prob (F-statistic):
0.00
Time:                             00:46:34   Log-Likelihood:
16824.
No. Observations:                     4625   AIC:
-3.364e+04
Df Residuals:                         4621   BIC:
-3.362e+04
Df Model:                                3
Covariance Type:                 nonrobust
==============================================================================
```

```
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const       8.948e-05      0.002      0.038      0.969      -0.004       0.005
EVR1           0.0481      0.003     19.069      0.000       0.043       0.053
EVR2           0.0347      0.006      6.003      0.000       0.023       0.046
EVR3          -0.0385      0.009     -4.118      0.000      -0.057      -0.020
==============================================================================
Omnibus:                     1750.358   Durbin-Watson:                   0.005
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             8057.429
Skew:                           1.794   Prob(JB):                         0.00
Kurtosis:                       8.379   Cond. No.                         117.
==============================================================================

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

```
Text(0.5, 1.0, 'EVR1 from Kendall correlation vs Credit Spreads')
```



All three correlation methods, Pearson, Spearman and Kendall, achieve a p-statistic = 0 for the EVR1 coefficient. Values of the R^2 statistic for Spearman and Kendall are better than Pearson's and more or less equal to each other. The standard error of the EVR1 coefficient in Spearman is slightly lower than Kendall's.

Correlation	EVR1 coefficient	R^2	EVR1 p-stat	EVR1 std err
Pearson	0.0364	0.296	0	0.003
Spearman	0.0389	0.346	0	0.002
Kendall	0.0481	0.349	0	0.003

This, and additionally the previous observation that historical data for EVR1 from Kendall correlation is much more conservative and as such detects less false positives (good) due to noise, but also true positives (bad) such as the 2008 financial crisis, leads us to choose the Spearman correlation as our methodology to extract the EVR1 signal.

Generate a Systemic Risk Indicator (SRI) from correlation breakdown and widening credit spreads An obvious observation that can be made of the above procedure that led us to pick the Spearman correlation matrix to generate the EVR1 metric, is that we applied linear regressions over a correlation measure that tracks nonlinearities.

The analysis would greatly benefit from employing Machine Learning in stead of linear regression to track these nonlinearities, at the price of course of more obfuscation in the selection process.

Indeed, the (Spearman EVR1 - credit spread) linear regression coefficient of 0.0389 is quite feeble, but this also might help justifying the construction of a SRI as linear combination of EVR1 and credit spread. The two terms are expected to correlate more heavily during crises, while a linear combination-based SRI would instead suggest EVR1 and credit spread to be independent from each other. However, a weak correlation when markets are placid might make an acceptable proxy for independence, while the linear combination will still cause the SRI to increase at market phase shifts, just more so than if the rising (EVR1-credit spreads) correlation were to be taken into account.

Therefore, we could define the SRI as following

$$\mathrm{SRI}(\text{credit spread}, \mathrm{EVR1}) = 10 \cdot \alpha \cdot \text{credit spread} + \beta \cdot \mathrm{EVR1}$$

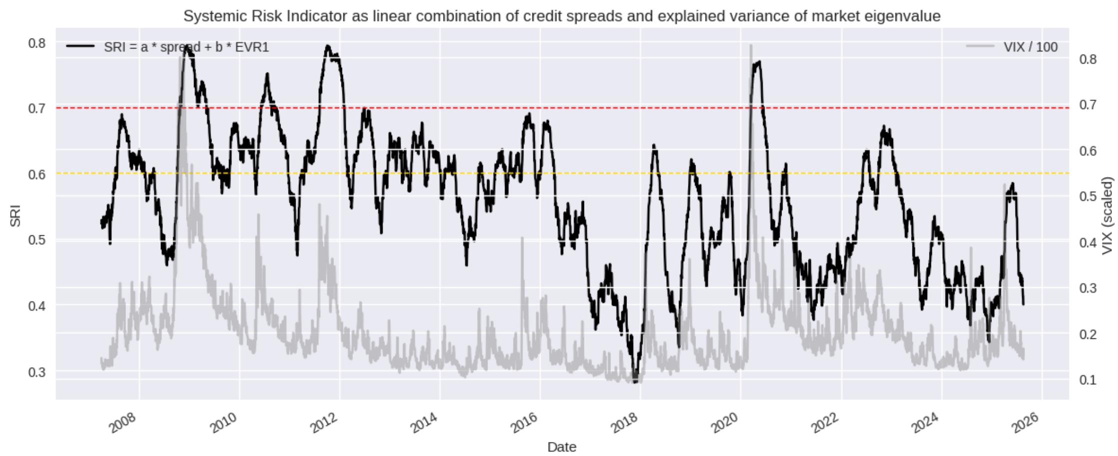
where coefficients α, β could be selected from a Least Squares procedure. The credit spread contribution is magnified tenfold as it is quantitatively an order of magnitude smaller than EVR1, and we want it to qualitatively matter.

The norm to minimise might be the distance of the SRI from the ^VIX index, as volatility increases during crises, and our goal is to generate an indicator of crises.

Assuming for now

$$\alpha = 0.2\beta = 0.8$$

We observe good correlation with ^VIX, but still the SRI signal for the 2008 financial crisis in the graph cannot be differentiated from false positives.



Student C: extension of a DCC-GARCH framework with funding stress indicators acting as regime change detectors Dynamical Conditional Correlation generalized autoregressive conditional heteroskedasticity (DCC-GARCH) is a model capable of estimating large time-varying covariance matrices. It does this by using estimating volatilities individually for each asset using GARCH for each separate time series.

The DCC (Dynamic Conditional Correlation) component is what makes the model "integrated" and powerful. It models how the correlations between these assets change over time. In a stable market, the correlation between stocks and bonds might be low or even negative. But during a market crisis, the correlation between almost all assets tends to spike towards 1 which is usually called a correlation spike.

In our case we will model use market sector tracker ETFs and market stress indicators as data for our correlation modeling, the idea is to be able to infer when the probability of correlation spikes are higher ahead of time. The data we are using are:

Sector tracking ETFs:

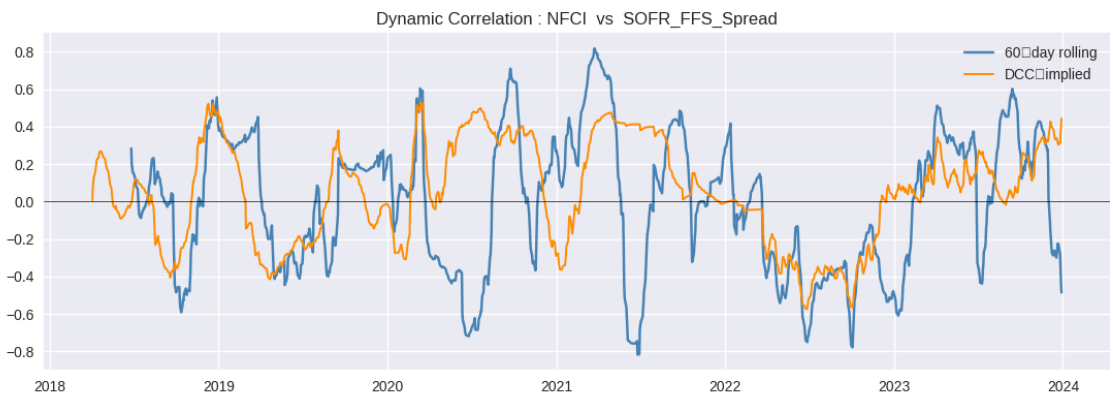
- XLY: Consumer Discretionary
- XLP: Consumer Staples
- XLE: Energy
- XLF: Financials
- XLV: Health Care
- XLI: Industrials
- XLB: Materials
- XLK: Technology
- XLU: Utilities #### Market stress indicators 1
- TED: Treasury-Eurodollar spread (This one has been discontinued by FRED in favour of Secured Overnight Financing Rate (SOFR))
- SOFR: Secured Overnight Financing Rate
- DFF: Federal Funds Effective Rate
- NFCI: Chicago Fed National Financial Conditions Index

- STLFSI4: St. Louis Fed Financial Stress Index
- VIX:CBOE Volatility Index

The difference between SOFR and DFF can be taken as a proxy of cross currency basis since it indicates dollar liquidity in the repo market. Adding these market indicators, builds on top of the materials provided by WQU.

After obtaining the data we run a $GARCH(1,1)$ model with a Gaussian error term. then we standardise the residuals,and collect the results in a dataframe.

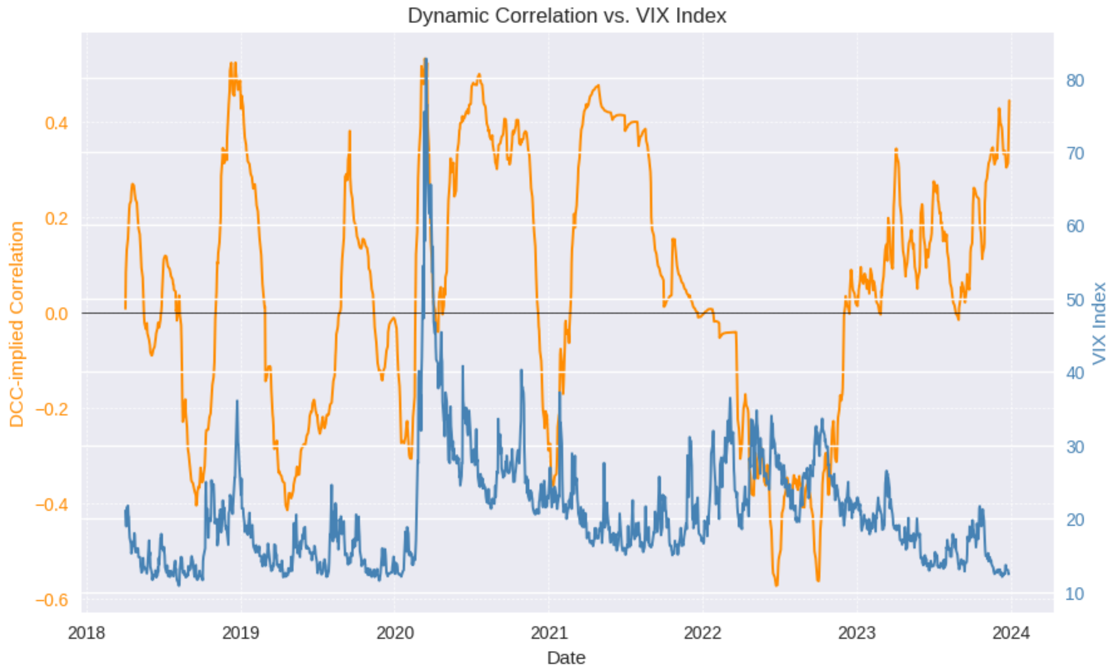
```
DCC  =0.050,   =0.900,  persistence=0.950
```



A cascade effect in financial markets describes a chain reaction where a shock to one part of the system rapidly spreads to others, creating a self-reinforcing loop of instability. This process often begins with a volatility shock—a sudden, sharp increase in market uncertainty, often triggered by an unexpected event.

A good proxy for this is the VIX which is a measure of the expected volatility of the market as a whole, when it is big the markets are quite volatile which usually means it is moving as a whole and correlations are close to one. We observe our DCC indicator (which measures the correlation between NFCI and our cross currency proxy) seems to be able to predict ahead of time when spikes in the vix will be present, we see that from 2023 the indicator breaks down, this is most likely due to the Fed changing its economic policies and entering into a high interest rate regime which had not been seen in many years 3, other correlation pairs were considered but this is the one we found gave the best predictions for low interest rates environments.

```
Text(0.5, 1.0, 'Dynamic Correlation vs. VIX Index')
```



This shock immediately leads to a correlation spike, as a sudden “risk-off” sentiment causes investors to sell all assets indiscriminately, driving their prices down in unison. Assets that were once uncorrelated or negatively correlated now move together, eliminating the benefits of diversification.

The increased volatility and correlation then trigger a funding freeze. As lenders become hesitant to extend credit due to heightened risk, market liquidity dries up. This forces institutions to sell assets to meet their funding needs, which in turn drives prices down further, causing even more volatility. This completes the feedback loop, as the funding freeze exacerbates the initial volatility shock. This cascade demonstrates how seemingly distinct market dynamics are deeply interconnected, creating a vicious cycle that can lead to a systemic crisis.

References

- FRED database, **https://fred.stlouisfed.org/series**, accessed on 25 August 2025