(EN) Public Report - Module 2

Student: Vinícius Oliveira Fernandes

Professor: Cristina Gramani

This project module aims to deepen the analysis of the unpredictability of economic recessions using historical data and the development of a Monte Carlo Simulation model.

The work was structured into five sprints:

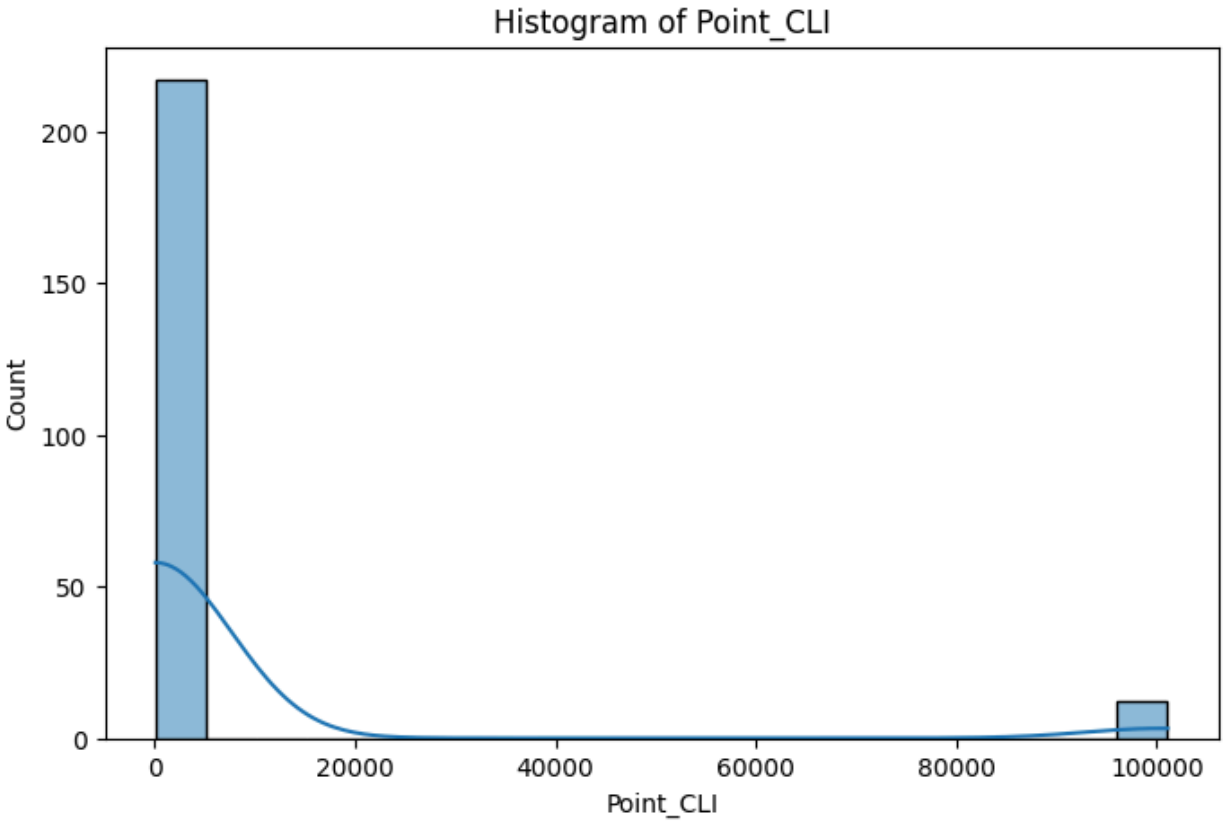
1. Collection and exploratory analysis of economic indicators
2. Initial implementation of the Monte Carlo Simulation
3. Refinement of parameters and simulation scenarios
4. Inference of the results obtained
5. Documentation of the insights generated

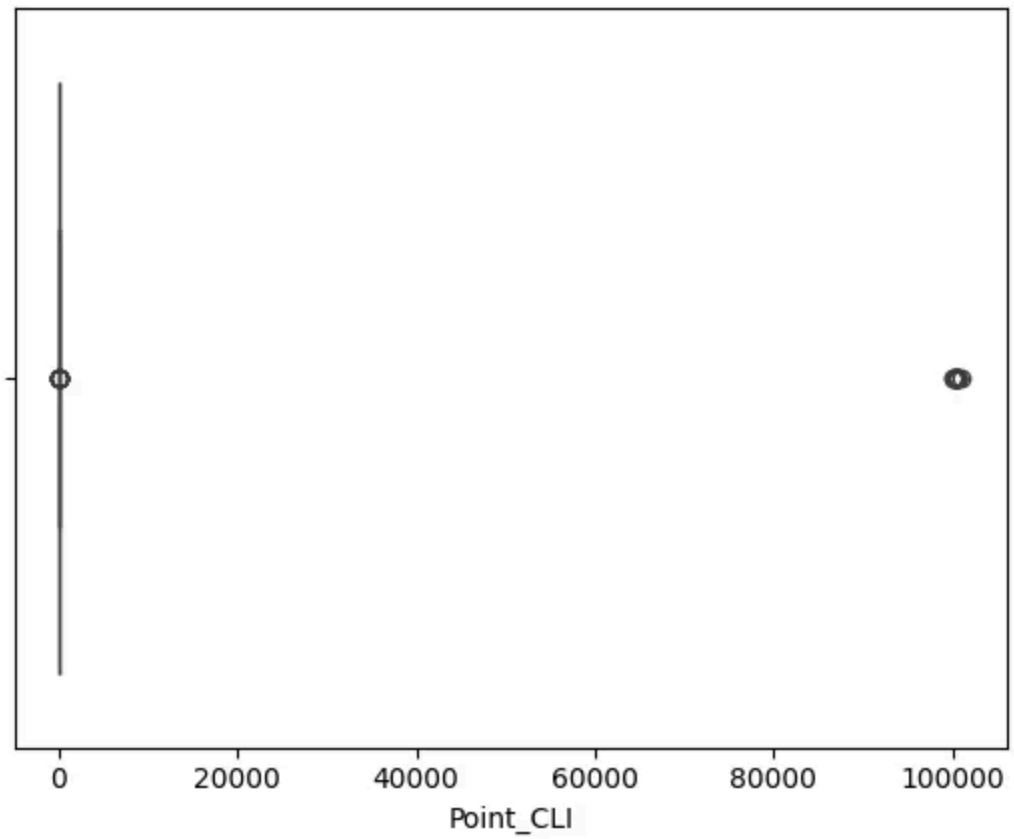
In this context, the development of the model and its premises was the main goal of the module. The central focus was to investigate whether these indicators have predictive capacity in the face of economic shocks, contributing to a more probabilistic interpretation of recessionary cycles.

Data Analysis

- **CLI**
  - Data from 2006 (*2006-03-01*) to 2025 (*2025-03-01*)
  - Statistical Summary:
    - Mean: 5357.988024
    - Std. Deviation: 22409.016195

- Min: 93.596980
- Quartiles:
  - 25% → 99.465290
  - 50% → 100.066300
  - 75% → 100.673800

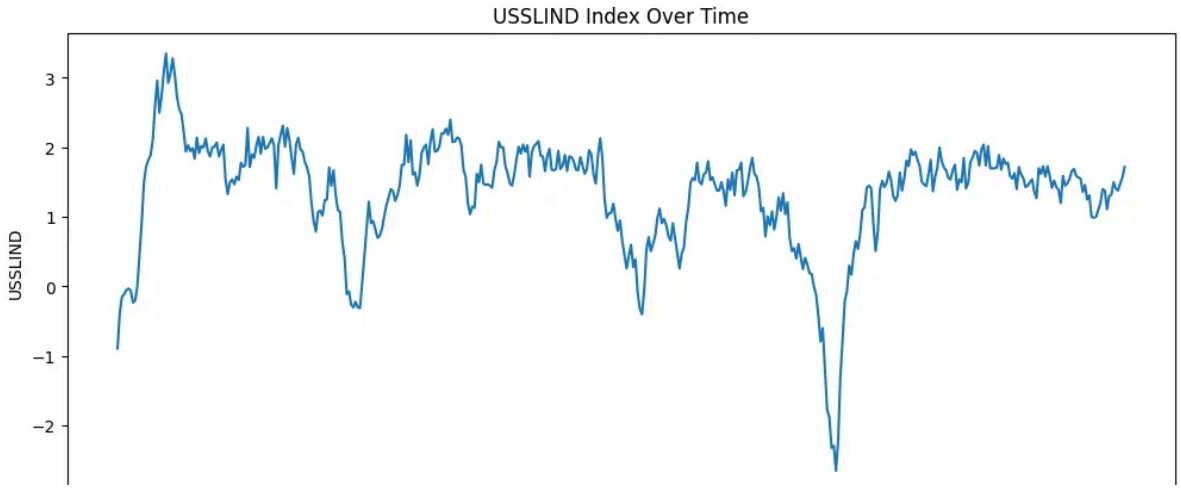


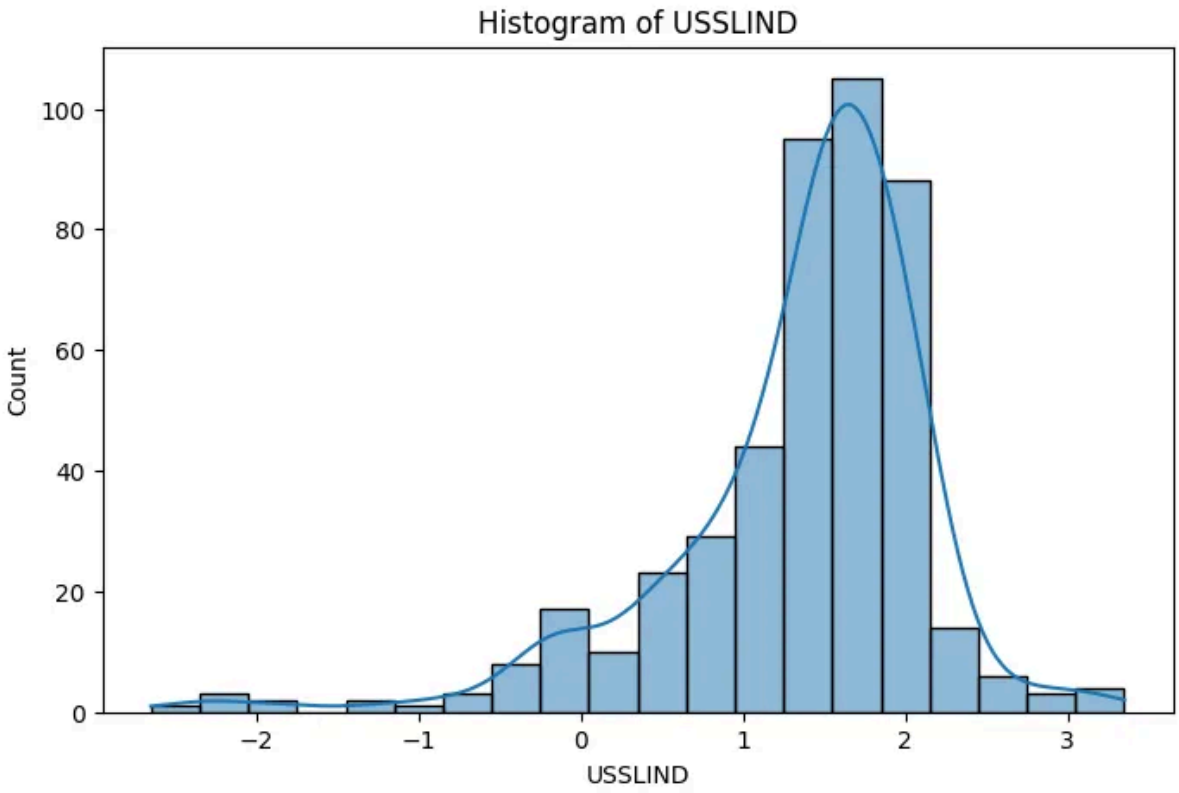


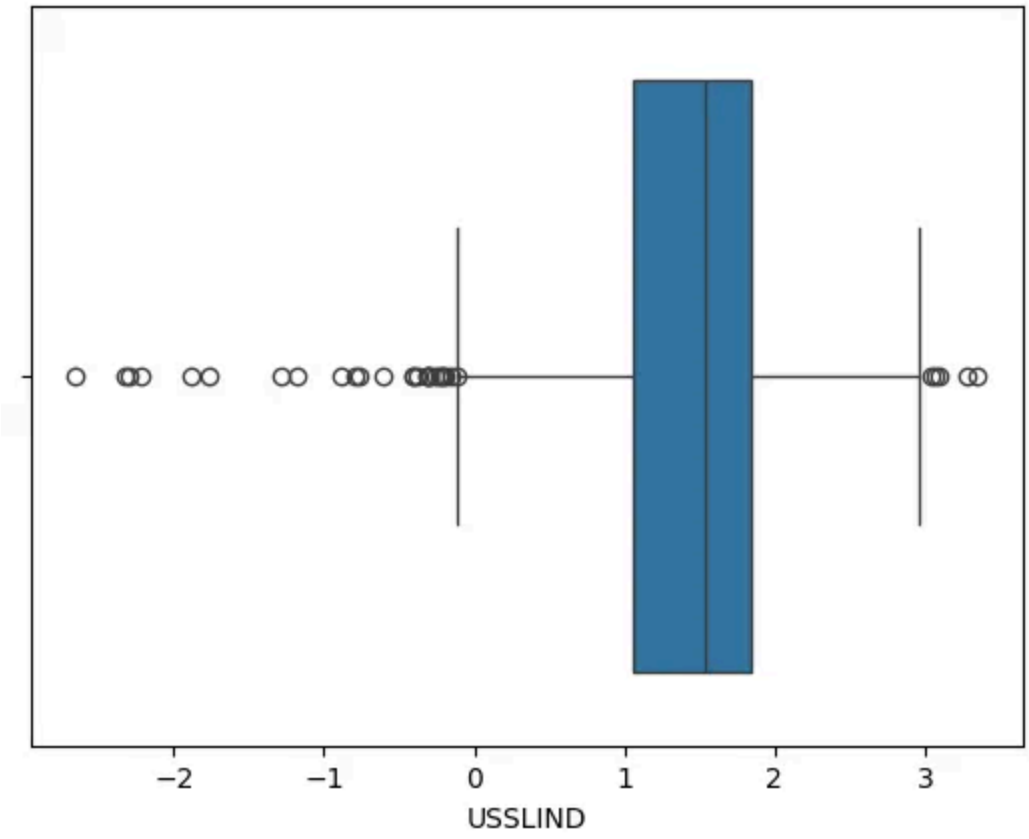
- The distribution is **left-skewed** (negative tail) but has a **concentrated peak between 1 and 2**, indicating frequently positive values.
    - A small number of negative values, considered outliers.
    - Several visible outliers, especially on the left side (negative values).
    - The median is close to 1.5.
- *Negative outliers are expected during times of crisis or recession.*

- **Federal Reserve LEI (USSLIND)**
    - Data from 1982 (*1982-01-01*) to 2020 (*2020-02-01*)
    - Statistical Summary:
        - Mean: 1.348974
        - Std. Deviation: 0.808302

- Min: -2.650000
- Quartiles:
  - 25% → 1.062500
  - 50% → 1.535000
  - 75% → 1.847500




Histogram of USSLIND
Count
USSLIND
0
20
40
60
80
100
−2
−1
0
1
2
3

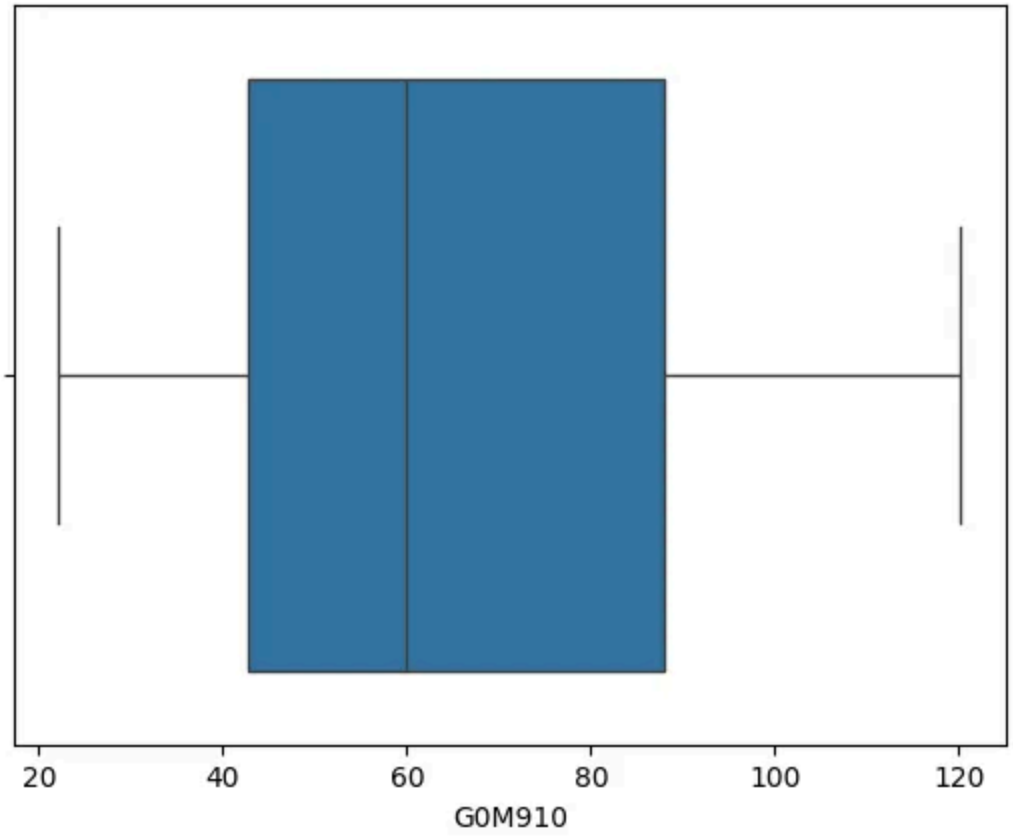

USSLIND
−2
−1
0
1
2
3

- Most of the CLI values are concentrated within a very narrow range (low values), suggesting a normal behavior.
    - There is **one extremely discrepant point**, with a value close to **100,000**, which is clearly a **severe outlier**.
        - This most likely represents a **data entry error** or a value that was **corrected/processed** prior to statistical analysis.

- **GDP per Capita (USA)**
    - Data from 1960 to 2023
    - Statistical Summary:
        - Mean: 29136.216664
        - Std. Deviation: 22026.012318
        - Min: 2999.864872
        - Quartiles:
            - 25% → 8394.554319
            - 50% → 24880.624841
            - 75% → 47408.769420

- **LEI – Conference Board (G0M910)**
    - Monthly data (795 observations)
    - Statistical Summary:
        - Mean: 66.288302
        - Std. Deviation: 26.935913
        - Min: 22.200000
        - Quartiles:
            - 25% → 42.800000

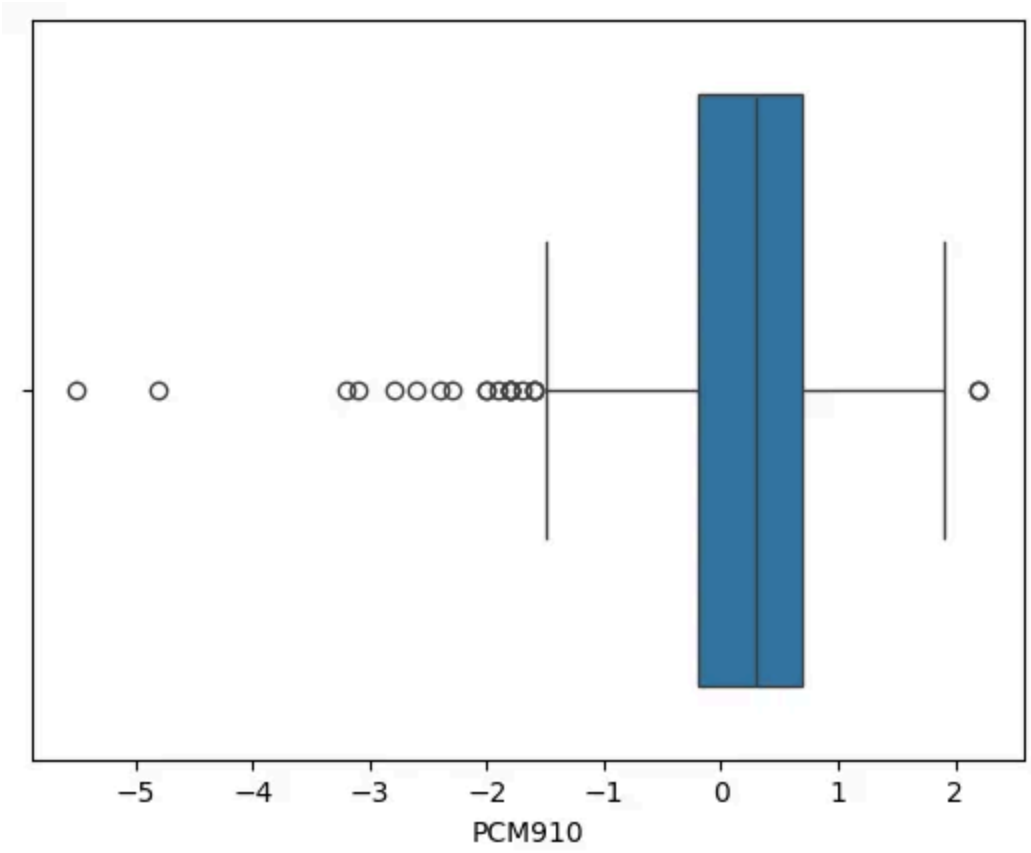
- 50% → 60.000000
- 75% → 88.050000



- **Symmetrical distribution**, with no visible outliers.
- Values range between 20 and 120, with a median close to 65.
- This variable is statistically well-behaved and can be used directly in predictive models without transformation.

- **LEI – Conference Board (PCM910)**
  - Monthly data (794 observations)
  - Statistical Summary:

- Mean: 0.192443
- Std. Deviation: 0.748630
- Min: -5.500000
- Quartiles:
  - 25% → -0.200000
  - 50% → 0.300000
  - 75% → 0.700000



- Asymmetric distribution with **several negative outliers** (up to -5%).
- The median is close to 0, suggesting the index changes very little on average from month to month.

- The outliers may correspond to abrupt economic events (market shocks, pandemics, crises).

Data Inference and Transformation

Indicator	Outliers?	Need to Transform/Normalize?	Notes
Point_CLI	Yes (very determinant to the analysis)	Yes, normalize	Outlier will impact analysis
USSLIND	Yes	Not necessary	Not impactful. Outliers are only specifically on certain dates
PCM910	Yes (little)	Not necessary	The variations are expected due to different periods of the years
G0M910	No	No	Stable and symmetric

All datasets were merged into a single DataFrame and normalized using MinMaxScaler (X′=(X−Xmin)/(Xmax−Xmin)(X' = (X - X_min) / (X_max - X_min)(X′=(X−Xmin)/(Xmax−Xmin). The only variable not normalized was GDP, since normalization significantly distorted the simulation results.

2008

24-Month Period

CLI

Data	CLI
2007-06-01 00:00:00	101.7681

2007-07-01 00:00:00	101.7095
2007-08-01 00:00:00	101.6182
2007-09-01 00:00:00	101.5185
2007-10-01 00:00:00	101.4225
2007-11-01 00:00:00	101.3274
2007-12-01 00:00:00	101.2353
2008-01-01 00:00:00	101.1377
2008-02-01 00:00:00	101.0243
2008-03-01 00:00:00	100.8861
2008-04-01 00:00:00	100.7035
2008-05-01 00:00:00	100448
2008-06-01 00:00:00	100.0939
2008-07-01 00:00:00	99.62764
2008-08-01 00:00:00	**99.04966**
2008-09-01 00:00:00	**98.37462**
2008-10-01 00:00:00	**97.64918**
2008-11-01 00:00:00	**96.9623**
2008-12-01 00:00:00	**96.39452**
2009-01-01 00:00:00	**96.00105**
2009-02-01 00:00:00	**95.81349**
2009-03-01 00:00:00	**95.84404**
2009-04-01 00:00:00	**96.07906**
2009-05-01 00:00:00	**96.46297**
2009-06-01 00:00:00	**96.93582**

From mid-2007 to mid-2009, the CLI begins to decline, with its lowest points (below the 25% quartile) concentrated in early to mid-2009—marking the “post-event” recession period when the economy was already fully suffering the consequences.

LEI - Conference Board

Date	G0M910	PCM910	P6M910
2007-06	97,1	-0,2	-3,5
2007-07	96,9	-0,2	-3,7
2007-08	96,5	-0,4	-5,3
2007-09	95,9	-0,6	-6,7
2007-10	95,5	-0,4	-8,5
2007-11	94,7	-0,8	-9,9
2007-12	**93,8**	**-1,0**	**-11,3**
2008-01	**92,7**	**-1,2**	**-11,6**
2008-02	**91,6**	**-1,2**	**-11,7**
2008-03	**90,3**	**-1,4**	**-10,8**
2008-04	**89,8**	**-0,6**	**-12,7**
2008-05	**89,0**	**-0,9**	**-12,7**
2008-06	**88,6**	**-0,4**	**-13,7**
2008-07	**86,6**	**-2,3**	**-18,2**
2008-08	**85,6**	**-1,2**	**-21,0**
2008-09	**83,9**	**-2,0**	**-24,7**
2008-10	**81,2**	**-3,2**	**-24,2**
2008-11	**79,1**	**-2,6**	**-24,5**
2008-12	76,9	-2,8	-23,9
2009-01	75,4	-2,0	-18,5
2009-02	74,4	-1,3	-13,2
2009-03	73,2	-1,6	-6,6
2009-04	73,3	0,1	-1,8
2009-05	73,7	0,5	2,7
2009-06	74,3	0,8	8,1

Code	Description

G0M910	Current level of the LEI
PCM910	Monthly percentage change
P6M910	Six-month cumulative change

The LEI indicators (G0M910, PCM910, and P6M910) begin showing negative variations as early as late 2007. The most severe drops in level and percentage change occurred from December 2007 through October 2008. These metrics clearly demonstrate the early signs of economic slowdown.

Date	G0M910	PCM910	P6M910
2007-12	93,8	-1,0	-11,3

The LEI is more sensitive than other indicators due to its responsiveness to sudden shocks. The percentage change indicators (PCM910 and P6M910) particularly stand out, showing signs of trouble months before the recession was officially declared.

US LEI Federal Reserve Bank

2007-06-01	0.70
2007-07-01	0.51
2007-08-01	0.55
2007-09-01	0.40
2007-10-01	0.61
2007-11-01	0.43
2007-12-01	0.25
2008-01-01	0.41
2008-02-01	**0.32**
2008-03-01	**0.19**
2008-04-01	**0.18**
2008-05-01	**0.00**
2008-06-01	**-0.12**
2008-07-01	**-0.41**
2008-08-01	**-0.79**

2008-09-01	**-0.60**
2008-10-01	**-1.18**
2008-11-01	**-1.77**
2008-12-01	**-1.88**
2009-01-01	**-2.32**
2009-02-01	**-2.29**
2009-03-01	-2.65
2009-04-01	-2.21
2009-05-01	-1.29
2009-06-01	-0.77

This index also predicted a downturn starting early in 2008, with a steep drop between March and December of that year. It is designed to forecast a state's economy over the next six months using a coincident index. Thus, it provided reasonably timely predictive value.

2020

Período de 12 meses

CLI

2020-01-01 00:00:00	99.1925
2020-02-01 00:00:00	99.17662
2020-03-01 00:00:00	**97.49301**
2020-04-01 00:00:00	**93.59698**
2020-05-01 00:00:00	95.07487
2020-06-01 00:00:00	97.06987
2020-07-01 00:00:00	98.29512
2020-08-01 00:00:00	98.77375

2020-09-01 00:00:00	98.97484
2020-10-01 00:00:00	99.21239
2020-11-01 00:00:00	99.47532
2020-12-01 00:00:00	99.8277
2021-01-01 00:00:00	100.0862

The index dropped significantly in March and April 2020 in response to the COVID-19 pandemic. April marked the lowest point, reflecting market panic. The index stayed below the historical 25% quartile until September, indicating a clear recessionary period.

LEI (Conference Board):

Date	G0M910	PCM910	P6M910
2020-02	112,9	0,0	-15,8
2020-03	107,5	-4,8	-11,6
2020-04	101,6	-5,5	-9,3
2020-05	103,5	1,9	-7,6
2020-06	105,8	2,2	3,0
2020-07	107,5	1,6	16,6
2020-08	108,5	0,9	13,8
2020-09	109,1	0,6	10,3
2020-10	109,7	0,5	8,5
2020-11	110,4	0,6	6,6
2020-12	111,1	0,6	7,5
2021-01	112,0	0,8	8,6
2021-02	112,0	0,0	8,7
2021-03	113,1	1,0	9,2

Similar patterns were observed in the LEI. Monthly and six-month changes reflected severe contraction during the height of the pandemic (March–May), capturing both the surprise and uncertainty of the crisis.

Defined Premise for the Simulations

Definition of what could be considered a recession:

- High percentage drop in the financial indicators defined as model inputs
    - Quartile analysis and selection of the bottom 0.25 quantile of each indicator to identify a trend shift
- "Flag" defined based on the premise derived from the historical data analysis of each indicator, as well as the behavior of these indicators during past recessions.

Purpose of the Monte Carlo Simulation: Simulate possible future paths for these indicators based on historical patterns.

- **Execution of the Monte Carlo Simulations →** For each scenario (in this case, each country), verify whether a recession occurs.
- **"Probability" of a recession →** Based on the simulation results, the probability of a recession is calculated.

Pre-simulation

Definition of thresholds

- Thresholds for each indicator are defined based on the **1st quartile (25%)** of their historical distribution.
- These values serve as **reference points to detect economically “low” levels** that may indicate a recession.

> Note: Although these thresholds are calculated, the is_recession function uses fixed values (cli < 100, g0 < 100, pcm < 0.25) instead of the thresholds;

> This was done because the thresholds produced poor results during simulation rounds.

Function to define recession markers:

```
def is_recession(cli, g0, pcm):
    return (
        cli < 100 and
        g0 < 100 and
        pcm < 0.25
    )
```

Monthly indicators

```
returns = df_merged[['Point_CLI', 'G0M910', 'PCM910']].pct_change().dropna()
```

The goal is to capture the **monthly percentage change (return)** of the three economic indicators.

- This represents the historical behavior of each variable over time.
- The distribution of these returns is used to simulate the future.

Simulation engine:

```
def monte_carlo_simulation(current_values, num_months, num_simulations=1000
```

- Each simulation starts with the current values of the indicators (`current_values`).
- The flag `recession_occurred` is initialized as `False`.

A further level of randomness is added—an **random shock** is applied each month:

- A return vector (`shock`) is randomly sampled from the historical return series (`returns`).
- This vector contains percentage changes for `Point_CLI`, `G0M910`, and `PCM910`.

- The indicator values are updated by multiplying by the shock factor:
  `new value = current value × (1 + simulated return)`
- This simulates the future behavior of the indicators based on historical volatility patterns.

Post fact:

- After each monthly shock is applied, the code checks whether the indicators **fall below the recession thresholds**.
- If this occurs in **any month** of the simulation, the `recession_occurred` flag is set to true.
  - Each simulation ends with `True` (if a recession occurred) or `False` (if not).
  - In the end, the return is a list of 100,000 boolean values.
  - Since `True` is treated as `1` and `False` as `0`, the **mean of the list** gives the **proportion of simulations** that indicated a recession.
  - This represents the **estimated probability** of a recession occurring within the next `n` months.

Reporte dos Resultados

United States - 2008 (SubPrime Crisis)

2008 Simulation Examples:

- **12-month forecast window: 61% probability of recession**
  - Historical window: 2007-06-12 to 2008-06-12

```python
def is_recession(cli, g0, pcm, uslin):
    return (
        cli < cli_threshold_for_2008 and
        g0 < g0_threshold_for_2008 and
        pcm < pcm_threshold_for_2008 and
        uslin < uslin_threshold_for_2008

    )
```
✓ 0.0s — Python

```python
returns_2008 = us_2008[['Point_CLI', 'G0M910', 'PCM910', 'USSLIND']].pct_change().dropna()

cli_returns_2008 = returns_2008[['Point_CLI']]
```
✓ 0.0s — Python

```python
current_values_2008 = us_2008[['Point_CLI', 'G0M910', 'PCM910', 'USSLIND']].iloc[-1].values

returns_2008 = monte_carlo_simulation(current_values_2008, 12)
prob_recession_2008 = np.mean(returns_2008)

print(f"Estimated probability of recession in next months: {prob_recession_2008:.2%}")
```
✓ 1m 11.9s — Python

```
/tmp/ipykernel_5270/4076599425.py:8: RuntimeWarning: invalid value encountered in multiply
  values *= (1 + shock)
Estimated probability of recession in next months: 61.02%
```

- **6-month forecast using only 2007 data: 89.32% probability of recession**
  - Historical window: full year of 2007

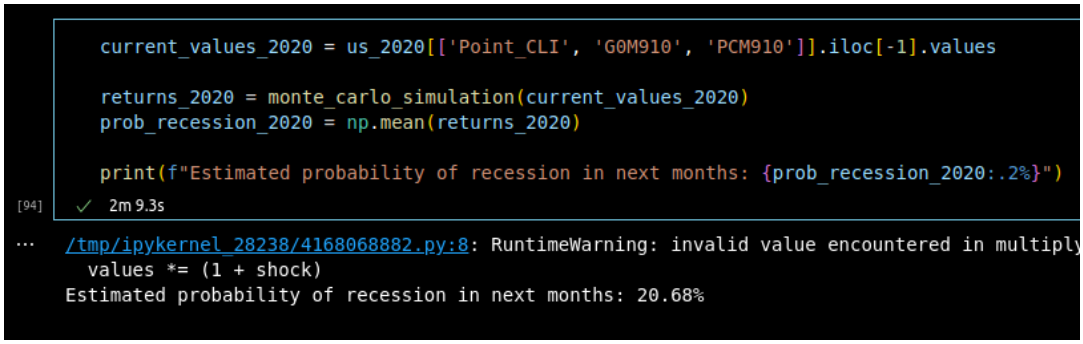
It was observed that using quartiles from smaller data windows yielded better simulation performance.

As a “bonus,” a broad-threshold test was run for the 2020 recession:

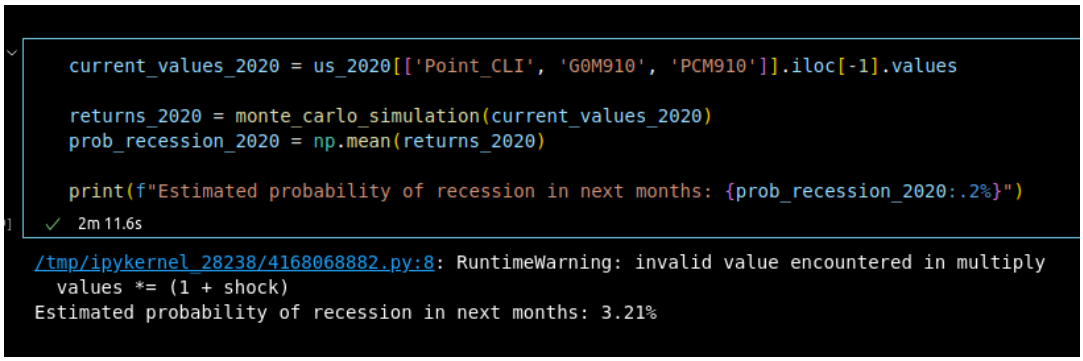
- **Result: 9% recession probability**

United States – 2020 (Pandemic)

- Monte Carlo Simulation Rounds: Different Assigned Time Windows
  - **6-Month Window:**
    - 2019-12 | 2020-01 | 2020-02 | 2020-03 | 2020-04 | 2020-05



  - **9-Month Window:**
    - **2019-09 | 2019-10 | 2019-11 | 2019-12 | 2020-01 | 2020-02 | 2020-03 | 2020-05**



  - **12-Month Window:**
    - 2019-07 | 2019-08 | 2019-09 | 2019-10 | 2019-11 | 2019-12 | 2020-01 | 2020-02 | 2020-03 | 2020-04 | 2020-05 | 2020-06

```
    current_values_2020 = us_2020[['Point_CLI', 'G0M910', 'PCM910']].iloc[-1].values

    returns_2020 = monte_carlo_simulation(current_values_2020)
    prob_recession_2020 = np.mean(returns_2020)

    print(f"Estimated probability of recession in next months: {prob_recession_2020:.2%}")
39] ✓ 2m 10.3s

/tmp/ipykernel_28238/4168068882.py:8: RuntimeWarning: invalid value encountered in multiply
  values *= (1 + shock)
Estimated probability of recession in next months: 1.47%
```

- **24-Month Window:**
  - 2019 and 2020

```
    current_values_2020 = us_2020[['Point_CLI', 'G0M910', 'PCM910']].iloc[-1].values

    returns_2020 = monte_carlo_simulation(current_values_2020)
    prob_recession_2020 = np.mean(returns_2020)

    print(f"Estimated probability of recession in next months: {prob_recession_2020:.2%}")
34] ✓ 2m 12.4s

/tmp/ipykernel_28238/4168068882.py:8: RuntimeWarning: invalid value encountered in multiply
  values *= (1 + shock)
Estimated probability of recession in next months: 0.07%
```