**3rd Graded Problem Set: "Extracting Predictive Equity Features from Stock Options"**

**Total Points:** 150

**Problem Statement**

Bakshi, Kapadia, Madan (2003, RFS, <https://doi.org/10.1093/rfs/16.1.0101>) showed how to derive the first four moments of the option-implied density in terms of the underlying option prices.

You obtain access to daily option implied volatility surface data, for February 2023, for both the SPX index and Goldman Sachs (GS).

**Task 1:** Compute the BKM (2003) first four moments of the underlying's return , , and , for the SPX index and all SP500 constituents on a daily basis for Feb 2023. Note, the file "sp500\_merged\_ivs\_2023-02.parquet" contains the daily IVS for all SP500 constituents. The file "spx\_ivs\_2023-02.parquet" contains the daily IVS for the SPX index. Hint: Definition of SKEW(t,tau) and KURT(t,tau) is given in eq (5) and (6) in BKM (2003).

**Task 2:** Decompose the BKM (2003) implied skewness (daily Feb 2023 for all SP500 constituents) into its sytematic and idiosyncratic part. Note, the file "sp500\_merged\_implBeta\_2023-02.parquet" contains the respective option implied betas of constituents. Hint: Skewness is the 3rd standardized moment. The 3rd cumulant is equal to the 3rd central moment of the distribution.

**Task 3:** BKM (2003) show in table 6 that the skewness of the index is more pronounced than that of single stocks. Can you confirm this observation based on the cross sectional data for Feb 2023? Give an economic interpretation/reasoning of your observation.

**Evaluation Criteria**

To achieve a perfect score, your submission must include:

1. A well-thought thru roadmap to answer Task 1, Task 2 and Task 3
2. **A well-structured, in-depth analysis** that is technically rigorous, clearly written, and intuitive.

**Submission Requirements**

Your submission must include:

* A **maximum** **five-page report** (11pt, including figures, tables, etc.).
* A **maximum of 12 presentation slides**, showcasing all relevant details.
  + **At least 5 slides must be non-Python-based** (motivation, theoretical explanations, visualizations, and insights).
  + **At least 3 slides must be Python-based**, demonstrating implementation and analysis.
* A **fully documented** Python script (or multiple scripts), clearly explaining your methodology and results.

**Final Notes**

* Ensure your report and presentation are **concise, well-organized, and professional**.
* Clearly justify all assumptions and methodological choices.
* Use high-quality visualizations to support your analysis.
* The Python code should be clean, efficient, and well-documented.
* We might pick a person by random to present the group’s solution

Good luck! 🚀