46-976 Financial Optimization, Project 2

For this assignment you may work in a group up to four students.

You may choose any of the four topics detailed below. You will need to make some calls on various details related to estimation, simulation, data, and portfolio construction. You are encouraged to be creative. Please feel free to incorporate techniques and data from any other courses or from your previous job experience.

Deliverables:

- Proposal (worth 3% and due by midnight on October 2): A one- or two-page document with the following information:

  1. Project title
  2. Team members
  3. Project's goals
  4. Description of the data that you will use
  5. Plan of action

  The main purpose of this proposal is to get you started on your project. I expect your reviewers (I will be one of them) to provide a default assessment "excellent" or "good" if your proposal shows that you have put some thought into what your team will do. Otherwise you may receive a "fair" or "poor" assessment. Please make sure your proposal describes the kind of data that you plan to use.

- Project highlights (worth 6% and due by midnight on October 15): A five-minute video (mp4 format), slide deck (pdf format), and demo code (zip format) that present the highlights of your project. The demo code should allow any of your peers to replicate the main results presented in your video and slide deck.

- Project write up (worth 3% and due by midnight on October 19): A report at most eight pages long with the following format:

  1. Executive summary
  2. Project description
  3. Main techniques/methods developed
  4. Results
  5. Conclusions

Peer reviews (worth 3%):

Each student in the class will serve as a peer reviewer for the first two deliverables (proposal, highlights) of a team different from his/hers. The reviewer is expected to provide an assessment and some brief constructive feedback in a professional and timely fashion. It is especially critical that each peer reviewer adheres to the following timeline:

1. Read the proposal of the team under review and provide a rough assessment (excellent/good/fair/poor) and feedback by the end of the day on Monday October 5, 2020.

2. Watch the video presentations and run the demo codes of the team under review and provide a rough assessment (excellent/good/fair/poor) and feedback by the end of the day on Sunday October 18, 2020.

These are some good sources of financial data (you are likely to know many more):

- Wharton Research Data Services (WRDS)
  `https://wrds.wharton.upenn.edu/wrdsauth/members.cgi`
- Kenneth French's page
  `http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`
- Yahoo finance
  `http://finance.yahoo.com`
- Google finance
  `http://www.google.com/finance`
- Bloomberg
  Bloomberg terminal and `http://www.bloomberg.com`
- MSCI Barra
  `http://www.mscibarra.com`

Each of the topics outlined fits the following format:

1. The goal of the project is to apply one or more of the techniques discussed in class.
2. You will first develop a *proof-of-concept* of the techniques via some artificially generated problem instance and data.
3. You will then develop an *actual implementation* the techniques in a more realistic context with an actual problem instance and actual data.

The steps in the outlined topics should be considered as general guidelines. Feel free to adapt them. That is, you may decide to omit, modify, or add some steps.

If you wish to pursue a topic that is completely different from those suggested in this document, please let me know by submitting your proposal directly to me by September 30 (two days before the official October 2 deadline). If the proposed project is interesting and comparable in terms of effort and merit to those suggested in this document, I will most likely agree to let your team work on it. If it is not, I will ask you to choose one of the suggested topics. (That is why you should submit your proposal a bit earlier.)

First Suggested Topic: Comparison of various covariance estimators

Goal: Compare various approaches to covariance estimation: sample covariance, single-factor, constant factor, and multiple-factor models.

Part I: Proof of concept

Suppose you have time-series data $\{\mathbf{r}(1), \mathbf{r}(2), \ldots, \mathbf{r}(T)\}$ for the vector of returns $\mathbf{r} \in \mathbb{R}^n$ of a universe of risky assets.

1. Develop an implementation of the following four estimators of the covariance matrix.
   (a) A naïve sample covariance matrix.
   (b) A covariance matrix based on a single-factor model
   $$\mathbf{r} = \boldsymbol{\beta} f + \mathbf{u},$$
   where either the loadings $\boldsymbol{\beta}$ or the single factor f are known. In the latter case the single factor f could be the returns of a benchmark portfolio.
   (c) A covariance matrix based on a constant correlation model. This assumes that the returns of every pair of different assets have the same correlation $\rho \in (-1, 1)$. This imposes the following structure on the covariance matrix $\mathbf{V}$:
   $$\mathbf{V}_{ii} = \sigma_i^2, \ \mathbf{V}_{ij} = \rho\sigma_i\sigma_j, \ \text{ for } \ i = 1, \ldots, n, \ \text{ and } \ j = 1, \ldots, n, \ \text{ with } \ i \neq j,$$
   or equivalently $\mathbf{V} = \rho\boldsymbol{\sigma}\boldsymbol{\sigma}^{\mathrm{T}} + (1-\rho)\mathrm{Diag}(\boldsymbol{\sigma})^2$.
   (d) A covariance matrix based on a multiple-factor model
   $$\mathbf{r} = \mathbf{B}\mathbf{f} + \mathbf{u}$$
   where the matrix of loadings $\mathbf{B}$ are known.
2. Construct artificial universes of n risky assets with (or possibly without) some linear factor structure. Generate artificial data $\{\mathbf{r}(1), \ldots, \mathbf{r}(T)\}$ and compare the performance of the fully-invested long-only minimum-variance portfolios obtained via the above estimation procedures. For a more informed comparison, evaluate the performance of the portfolios in-sample and out-of-sample. You could also compare the performance of other portfolios such as fully-invested maximum-diversification portfolio and/or the fully invested risk-parity portfolio.

   There are a number of important details that you should explore and make choices about: the size n of the universe, the linear factor structure, the number of factors, etc.

Part II: Actual implementation

Select a universe of at least 20 or more risky assets (e.g., stocks) and some potential factors for which you can compute exposures. The factors could be industries (energy, financials, etc), style factors (value, momentum, etc). Collect data on their returns over some time horizon.

Perform the estimation of covariance matrix along the lines of the above proof-of-concept. Evaluate your results both in-sample and out-of-sample.

Second Suggested Topic: Comparison of various diversification approaches

Goal: Compare various approaches to diversification: minimum-variance, risk-parity, hierarchical risk-parity

Part I: Proof of concept

Suppose you have the covariance matrix $\mathbf{V} = \text{cov}(\mathbf{r})$ of the vector of returns $\mathbf{r} \in \mathbb{R}^n$ of a universe of risky assets.

1. Implement an optimization model to compute each of the following four types of diversified portfolios. In all cases assume you are constructing a fully-invested portfolio and include flexibility in your model so that you can turn on or tune the following two constraints: long-only constraint and a constraint to limit the component-wise deviation from the equally-weighted benchmark portfolio.

   (a) Minimum-variance portfolio

   (b) Risk-parity portfolio

   (c) Maximum-diversification portfolio

   (d) Hierarchical risk-parity

2. Choose a positive definite covariance matrix $\mathbf{V}$ and construct an artificial universe of n risky assets whose returns $\mathbf{r}$ satisfy $\mathbf{V} = \text{cov}(\mathbf{r})$ and $\mathbb{E}(\mathbf{r})$ is a multiple of $\mathbf{1}$, i.e., all expected returns are the same.

   Compute the above diversified portfolios for the covariance matrix $\mathbf{V}$. Compare the performance of these portfolios and the equally weighted portfolio on some artificially generated data $\{\mathbf{r}(1), \ldots, \mathbf{r}(T)\}$. As "default options" include the long-only constraint but no component-wise constraints in your models.

   Examine how your results change when you play with the following issues:

   - Covariance matrices of different characteristics: no correlation (i.e., diagonal), high correlation, both low and high volatility assets.
   - Omit the long-only constraint and/or include a constraint to limit the component-wise deviations from the benchmark constraint.
   - Suppose the above optimization models do not work with the exact matrix $\mathbf{V}$ but instead with an estimate $\hat{\mathbf{V}}$ of $\mathbf{V}$ that includes some kind of estimation error.

Part II: Actual implementation

Select a universe of at least 20 or more risky assets (e.g., stocks). Collect data on their returns over some time horizon and choose a procedure to estimate the covariance matrix $\mathbf{V}$.

Compute the above types of diversified portfolios. Compare their performance along lines of the above proof-of-concept both in-sample and out-of-sample.

Third Suggested Topic: Construction of an index fund

Goal: Construct a small fund to track a market index.

Part I: Proof of concept

Suppose you have a universe of n risky stocks with covariance of returns $\mathbf{V}$. You want to construct a portfolio that includes at most $k \ll n$ stocks and that tracks a specific benchmark portfolio as closely as possible.

1. Implement a model that performs the following steps:

    - Use some kind of clustering or variable selection approach to select the set of k assets that will be included in the tracking portfolio. A possible criterion for clustering is a distance based on the correlation matrix.
    - Assign portfolio weights to the selected assets. A possible approach is to mimic the the weighting method used in the index. For example, for a market-value weighted index, assign the weight of each selected stock according to the market value of all the stocks that it represents. You can also attempt to replicate various attributes of the benchmark portfolio, such as beta (relative to the benchmark) equal to one, exposure to particular sectors (subsets of the universe), etc.
    - Include some flexibility in your model so that you can easily incorporate a turnover constraint.

2. Choose a positive definite covariance matrix $\mathbf{V}$ and construct an artificial universe of n risky assets whose returns $\mathbf{r}$ satisfy $\mathbf{V} = \mathrm{cov}(\mathbf{r})$ and $\mathbb{E}(\mathbf{r})$ is a multiple of $\mathbf{1}$, i.e., all expected returns are the same. Consider the following two benchmarks. First, the equally-weighted portfolio with holdings
$$x_i = \frac{1}{n},\ i = 1, \ldots, n.$$
Second, the value-weighted portfolio that uses the inverse of volatility as a proxy for size:
$$x_i = \frac{1/\sigma_i}{\sum_{j=1}^n 1/\sigma_j},\ i = 1, \ldots, n$$
where σ_i denotes the standard deviation (volatility) of the return of asset i.

    Compute a tracking portfolio for each of the two benchmarks for various values of $k \ll n$. Compare the performance of these portfolios and of the benchmark portfolio on some artificially generated data $\{\mathbf{r}(1), \ldots, \mathbf{r}(T)\}$. Report your results in terms of tracking error and turnover. Examine how your results vary if you play with the rebalancing frequency and with an upper-bound constraint on the amount of turnover.

Part II: Actual implementation

Select a universe of at least $n = 25$ or more risky assets (e.g., stocks). Collect data on their returns over some time horizon and choose a procedure to estimate the covariance matrix $\mathbf{V}$. Choose also a suitable benchmark, e.g., an equally-weighted or a value-weighted portfolio.

Compute portfolios to track the benchmark with various values of k. Compare their performance along lines of the above proof-of-concept both in-sample and out-of-sample.

Fourth Suggested Topic: Implementation of the Black-Litterman model

Goal: Test the effectiveness of the Black-Litterman model to incorporate views on the expected returns of a universe of risky assets.

Part I: Proof of concept

Suppose you have a universe of n risky stocks with covariance of returns $\mathbf{V} \in \mathbb{R}^{n\times n}$, a vector of expected excess returns $\mu \in \mathbb{R}^n$, and a fully-invested long-only benchmark portfolio $\mathbf{x}_B \in \mathbb{R}^n$.

Assume the following are known: The covariance matrix $\mathbf{V} \in \mathbb{R}^{n\times n}$, the consensus (equilibrium, or CAPM) excess returns $\boldsymbol{\pi} \in \mathbb{R}^n$ implied by the benchmark portfolio, and some set of views on the vector of excess expected returns $\boldsymbol{\mu}$ stated as

$$\mathbf{P}\boldsymbol{\mu} = \mathbf{q}$$

for some $\mathbf{P} \in \mathbb{R}^{k\times n}, \mathbf{q} \in \mathbb{R}^k$ where $k < n$. For consistency in the model, assume that indeed the (unknown) vector of expected returns $\boldsymbol{\mu}$ satisfies $\mathbf{P}\boldsymbol{\mu} = \mathbf{q}$ and assume also that the consensus returns satisfy the CAPM equation

$$\boldsymbol{\pi} = \boldsymbol{\beta}\mu_B$$

where $\mu_B = \boldsymbol{\mu}^{\mathrm{T}}\mathbf{x}_B$ and $\boldsymbol{\beta} = \frac{1}{\mathbf{x}_B^{\mathrm{T}}\mathbf{V}\mathbf{x}_B}\mathbf{V}\mathbf{x}_B$.

1. Implement a portfolio optimization model that takes the above inputs and performs the following steps:
   - Compute the posterior returns $\hat{\boldsymbol{\mu}}$ that incorporate the views $\mathbf{P}\boldsymbol{\mu} = \mathbf{q}$ according to the Black-Litterman model.
   - Choose a new fully-invested long-only optimal portfolio intended to have better characteristics than the benchmark portfolio $\mathbf{x}_B$. You have several possible choices to set up this portfolio construction model. A simple and straightforward approach is to compute the fully-invested long-only portfolio that maximizes the Sharpe ratio for the posterior expected returns $\hat{\boldsymbol{\mu}}$. However, a more interesting approach is to look for the portfolio with maximum expected return subject to some combination of the following constraints: satisfy a constraint on the component-wise deviations from the benchmark, satisfy a upper bound constraint on the tracking error, satisfy an upper bound constraint on the portfolio variance, etc.
   - Include some flexibility in your model so that the constraints can be easily modified, tuned, or omitted altogether.

2. Choose a positive definite covariance matrix $\mathbf{V} \in \mathbb{R}^n$, a vector of expected excess returns $\boldsymbol{\mu} \in \mathbb{R}^n$, and a fully-invested long-only benchmark portfolio $\mathbf{x}_B \in \mathbb{R}^n$. Compute the equilibrium expected return vector $\boldsymbol{\pi}$ and select a set of views $\mathbf{P}\boldsymbol{\mu} = \mathbf{q}$.

   Compare the performance of the benchmark portfolio and of the Black-Litterman portfolios on some artificially generated data $\{\mathbf{r}(1), \ldots, \mathbf{r}(T)\}$. Examine how your results vary if you play with the number of views (that is, the number of rows k in $\mathbf{P}$) and if you also play with the constraints included in your portfolio construction model.

Part II: Actual implementation

Select a universe of risky assets. You may choose a handful of asset classes or a larger set of assets (e.g., stocks). It may be more manageable to work with an instance of the former type, i.e., an asset allocation problem. Choose also a benchmark portfolio. Two possible and natural choices are the equally-weighted portfolio and the value-weighted portfolio.

Collect data on the excess returns of the risky assets over some time horizon and choose a procedure to estimate the covariance matrix $\mathbf{V}$ and equilibrium returns $\boldsymbol{\pi}$. Construct views on your assets via the following kind of *cheating procedure:* suppose you had perfect foresight on the expected return q of a portfolio $\mathbf{p}$. You could use that foresight to construct the view $\mathbf{p}^{\mathrm{T}}\boldsymbol{\mu} = q$. A set of views would correspond to this kind of foresight for various portfolios.

Compute the Black-Litterman portfolios obtained via the implementation developed in the above proof-of-concept. Compare their performance with the performance of the benchmark portfolio along the along lines of the above proof-of-concept both in-sample and out-of-sample.