Financial Analytics
Martin Haugh

Imperial College Business School

Assignment 3

Due: 11.59pm Monday 24^{th} May 2021

Rules

1. This is a group assignment. (There are approximately 3 people per group and by now you should know your assigned group.)
2. You are free to use `R` or `Python` for the programming components of this assignment.
3. Within each group **I strongly encourage each person to attempt each question by his / herself first** before discussing it with other members of the group.
4. Students should **not** consult students in other groups when working on their assignments.
5. Late assignments will **not** be accepted and all assignments must be submitted through the Hub with one assignment submission per group. Your submission should include a PDF report with your answers to each question together with screenshots of any relevant code. Make sure your PDF clearly identifies each member of the group by CID and name.

The **first two** questions require some familiarity with the `Tidyverse` suite of packages in `R` or the `Pandas` library in `Python`. Every student doing an MSc in Business Analytics should attempt to build some basic level of competence with either the `Tidyverse` or `Pandas` as these packages are the key tools that are required to "*slice-and-dice*" data. Note that these questions must be done programmatically in `R` or `Python` and not using `Excel`'s pivot-table functionality. (It's fine if you want to use `Excel`'s pivot-table functionality to check some of your results but the code screenshots that you submit must be in `R` or `Python`.)

Remark: If you choose to take the `R-Tidyverse` approach then an excellent online resource is *R for Data Science* by Grolemund and Wickham. Everything you need can be found here. You may also want to consult the excellent *RStudio* cheat-sheets that relate to the `Tidyverse`. The `read_csv` function is particularly handy for reading in data from csv files! Most of what you need, however, can be found in the *Tidyverse_Intro* PDF together with the corresponding `R Notebook` *DataWrangling_NYCFlights13.Rmd* which have been posted on the course web-site.

---

1. **Scenario Analysis I (30 marks)**
   Load the data from the *StressData.csv* file. Each row of the file contains data on a particular security (option or futures contract) that arises from a particular scenario. The scenarios that were considered are all combinations of the underlying moving by $\{\pm 20\%, \pm 10\%, \pm 5\%, \pm 2\%, \pm 1\%, 0\%\}$ and implied volatilities moving by $\{\pm 10, \pm 5, \pm 2, \pm 1, 0\}$ percentage points. There are therefore a total of $11 \times 9 = 99$ scenarios. Since there are a total of 83 securities (see the *Options-Data.csv* file referred to in Exercise 2 below) this results in a total (excluding the headings

row) of $99 \times 83 = 8,217$ rows in the *StressData.csv* file. The goal of this question is to do some slicing-and-dicing of the data in the *StressData.csv* file. As is often the case, most of the variables, i.e. columns, are irrelevant but I've left them in the file because (a) that's how data files often appear in practice and (b) you may want to play around with some of the other variables at a later time!

Construct a pivot-table showing the total P&L in each scenario for all securities with the S&P 500 as the underlying security. Your table should be laid out like the table in Figure 9 of Section 4.2 of the *An Introduction to Derivatives Pricing* lecture notes. In particular, your results should be displayed as an 11×9 matrix. (The columns / variables you will need for this are 'Underlying Stress', 'Volatility Stress', 'Underlying' and 'PnL'.)

Solution: See the `R Notebook` *ScenarioAnalysis_for_OptionsPortfolios.Rmd*.

---

2. **Delta-Vega-Gamma Approximations and Scenario Analysis II (30 marks)**
Load the data from the *OptionsData.csv* file. Excluding the header row, there are 83 rows in this file - one for each security.

   (a) Write a piece of code to display the sum of the 'Total \$Delta' (column V) grouped by the underlying security. Since there are just three underlying securities, you should output three 'Total \$Delta's: one for each underlying. (Note that each of these "Total \$Delta"s are denominated in different currencies so it wouldn't make sense to add them together without first converting them all to a common currency. Note also that it's easy to check your work within the *OptionsData.csv* file using `Excel`'s filtering functionality together with the SUBTOTAL function. But you'd need to use the right arguments with the SUBTOTAL function!)

   Are these numbers consistent (in the case of the SPX) with your pivot-table from Exercise 1? **(10 marks)**

   Solution: See the `R Notebook` *ScenarioAnalysis_for_OptionsPortfolios.Rmd*. The 'Total \$Delta's are -164,156 (yen), 24,442 (USD) and -17,486 (Euro) for the NKY, SPX and SX5E, respectively.

   And yes, these numbers (in the case of the SPX) are consistent with the pivot-table from Exercise 4. A total-dollar-delta of 24,442 suggests that if the SPX were to rise by say 1% and implied volatilities remain unchanged, then the total P&L in the portfolio should be *approximately* $1\% \times 24,442 \approx 244$. This is very close to the true number of 233 in the pivot-table. Similarly a fall of 2% in the SPX would suggest a total P&L of approx $-2\% \times 24,422 \approx -488$ which is reasonably close to the true value of -531 from the pivot-table.

   (b) Repeat part (a) but this time compute the sum of 'Total \$Gamma' by underlying security. Again in the case of the SPX, are your numbers consistent with the results in your pivot-table from Exercise 1? **(10 marks)**

Solution: See the `R Notebook` *ScenarioAnalysis_for_OptionsPortfolios.Rmd*. The sum of the "Total \$Gamma"'s are -1,508,038, -103,722 and -458,146 for the NKY, SPX and SX5E, respectively.

And yes, they are also consistent with the results from the pivot-table in Exercise 1. For example, if the SPX increases by +10% then the predicted P&L from the 'Total \$Delta' and 'Total \$Gamma' is

$$\begin{aligned} \text{Predicted P\&L} &= 10\% \times 24,442 + (10\%)^2 \times -103,722 \\ &\approx 2,444 - 1,037 \\ &= 1,407 \end{aligned}$$

which is quite close to the true value of 1,483 reported in the pivot-table.

(c) Finally, repeat part (a) for 'Total Vega 1%'. Again in the case of the SPX, are your numbers consistent with the results in your pivot-table from Exercise 1? **(10 marks)**

Solution: See the `R Notebook` *ScenarioAnalysis_for_OptionsPortfolios.Rmd*. It turns out that the sum of the 'Total Vega 1%'s are -13,304, -1,074 and -3,809 for the NKY, SPX and SX5E, respectively. And yes, the reported value of -1,074 for the SPX is consistent with the pivot-table for we see there that an increase of 1% in implied volatilities results in a P&L of -1,071 which is very close indeed to -1,074.

---

3. **Simulation and Construction of Actual, Estimated and True Efficient Frontiers (40 marks)**

The purpose of this exercise[1] is to explore the effect of estimation error on the computation of efficient portfolios by comparing the "true", "estimated", and "actual" efficient frontiers. To that end, assume the expected return and covariance matrix in the `Excel` spreadsheet *Return-Covariance-Data.xlsx* are the "true" values for the expected returns and covariances for a set of eight assets. These are *monthly* expected returns and covariances.

Next, using these "true" values and assuming a multivariate normal for the returns, generate monthly returns for 5 years. (You may find the *mvrnorm* function in `R` useful for doing this.)

(a) Compute the sample mean and the sample covariance matrix of the returns you generated. **(5 marks)**

(b) Compute at least ten long-only efficient portfolios along the efficient frontier based on the estimates you found in part (a). Choose efficient portfolios whose expected returns range from that of the long-only minimum-variance portfolio to that of the long-only portfolio with maximum expected returns. Save these efficient portfolios. (To be clear, all of these calculations are done using the sample mean and the sample covariance matrix from part (a).) **(12 marks)**

[1] This exercise is taken from *Optimization Methods in Finance* (2^{nd} edition) by Cornuéjols, Peña and Tütüncü and published by Cambridge University Press.

(c) Now compute the "actual" expected returns and standard deviations for the portfolios found in step (b). These are the values of true expected returns and standard deviations of these portfolios. **(6 marks)**

(d) On the same figure plot the "estimated" efficient frontier found in (b), the "actual" frontier from step (c), and the "true" frontier, i.e. the one we would get if we used the true parameters. (When constructing the true frontier it will be easiest to do so with expected returns ranging from that of the *true* long-only minimum-variance portfolio to that of the long-only portfolio with maximum *true* expected returns.) **(12 marks)**

(e) Repeat the above steps (generate a 5-year history, estimate, compute efficient portfolios) a few times. What do you observe? **(5 marks)**

Solution: See the `R Notebook` *Solution_Actual-Est-True-Frontiers.Rmd*.
