Washington University — Fall 2009
Department of Economics — James Morley

Economics 487

Project

Proposal due Tuesday 11/10
Final Project due Wednesday 12/9 (by 5:00pm)
(20% penalty per day if the project is submitted late)

For this project, you will analyze the behaviour of 10 stocks listed on the NYSE or the Nasdaq using Excel to make calculations and answer questions. The project has two stages. The first stage is the proposal, for which you will choose the stocks, make basic return calculations, calculate summary statistics, and write up a short discussion about your data. In the second stage, you will make formal portfolio theory calculations, market model calculations, CAPM calculations, test the assumptions of your models, and write up a formal term paper following the guidelines set out below.

Project Tasks:
1. Downloading data
2. Return Calculations
3. Summary Statistics
4. Portfolio Theory Calculations
5. Market Model Calculations
6. CAPM Calculations
7. Testing Assumptions

1. Downloading Data

Collect price data for 10 stocks listed on the NYSE or Nasdaq exchanges over the period from December 1994 to the present. Try to have your stocks represent at least a few very different industries with what you would expect to be different levels of risk. You will also collect price data for a broad market index (e.g., the S&P500) and yield data for a short-term T-bill. Include small stocks as well as large stocks to make sure your portfolio variance is similar to that of the market.

a. Go to www.yahoo.com and click on the *Finance* link.
b. In the "Enter Symbol" box at the top, type the symbol of the stock for which you want data. For example, the symbol for Microsoft in MSFT. If you do not know the symbol for a company, select the *symbol lookup* link and type in the name of the company.
c. Record the current market capitalization ("Market Cap").
d. On the left, under *Quotes*, click on *Historical Prices*.
e. In the SET DATE RANGE box, select *monthly* and change the start date to December 1, 1994. Click *Get Prices*.
f. This will bring up the table with price data for the selected time period. Make sure that the data are available at the beginning of the sample period (you will have to choose a different company if it does not have a long enough history of share prices).

g. At the bottom of the screen, select *Download to Spreadsheet*. Change the name of the file to the symbol name (e.g., msft.csv). Note: csv stands for comma separated value file, which is easily read by Excel.
h. Open the Excel file. The file will have the price data in descending order (most recent data is at the top). You will want the data in ascending order. To do this, highlight all of the data (including column headers). Then, select Data/Sort, which brings up the Sort dialogue box. The default should have sort by date with ascending order selected. If so, click OK. Otherwise, select those options and then click OK.
i. Save the file. Note that the data you will use in the project are the *Adj. Close* data.
j. When you have done this for 10 stocks, do the same for the market index. The symbol for the S&P 500 is "^gspc". You can also find it by selecting the *Indices* link from the initial Yahoo! Financial Quotes page. Name the file "sp500.cvs". (As with individual stocks, make sure to use adjusted price measure that includes dividends.)
k. Finally, you will need data for the risk-free asset. Again, you can get data from the *Indices* link. You want yield data for the 13-week bill under the Treasury Securities heading. The symbol is "^irx". For this series, you only need data from January 1, 1995 to the present. Name the file "tbill.cvs".

2. Return Calculations

The next thing to do is to calculate continuously compounded returns for the ten stocks, the market index, and the risk-free rate. For the price data, the procedure is standard. For the risk-free rate, the given data is not price data, but annualized percentage yields. To convert to a continuously compounded return, you need to first divide by 100 to get the yield as a decimal. Then, take the natural log of (1+yield). Finally, you can convert to a monthly continuously compounded return by dividing by 12. Also, construct an equal-weighted portfolio for your 10 stocks (i.e., construct a return series that is an equal-weighted average of the returns for the 10 stocks) and a value-weighted portfolio for your 10 stocks (i.e., construct a return series that is a weighted average of the returns for the 10 stocks, where the weights are proportional to the current Market Cap for each firm). For the value-weighted portfolio, a specific firm's implicit weight is its market cap divided by the sum of the market caps for all ten firms (collect all ten market caps on the same day). Put all of the return data in new Excel File.

3. Summary Statistics

a. Compute time plots of each of the 12 return series (the 10 individual stock returns and the equal-weighted and value-weighted returns). When reporting plots for each return, report along with S&P500 return and risk-free return on each graph. Also, report the Market Cap numbers for each firm and the corresponding weights used in constructing the value-weighted return series.
b. Compute histograms for each of the 14 series (the 10 individual stock returns, the equal- and value-weighted returns, the S&P500 return, and the risk-free return). Do they look Normal?
c. Compute mean, variance, standard deviation, skewness, and kurtosis for each of the 14 series. Report kurtosis (not excess kurtosis measured by Excel). If the mean is negative for any of the series, compute the median. Is the median positive? If so, why do you think there is a difference between the mean and median? Also, compare the

standard deviations of the other 13 series to the standard deviation of the S&P500 return series.

d. Compute the 95% confidence intervals for the estimates of the means and standard deviations of each of the 14 series. (Use the formulas in "hw2n.pdf" on the class website.)
e. Compute the (10x10) sample covariance matrix for 10 return series for the individual stocks. Use the covariance tool under Data Analysis.
f. Compute the Sharpe Ratio for the 10 return series for the individual stocks and for the S&P500. You can use the average t-bill yield to proxy for the risk-free rate in these calculations. This is equivalent to using the mean excess return for each stock to estimate its risk premium. For the mean returns, use the median estimate if the excess mean based on the sample average is negative. If it is still negative for the median estimate, set the excess mean to zero.

PROJECT PROPOSAL

Having done these first three tasks, you are ready to write up your project proposal. In the proposal, list the stocks you chose to examine, with some discussion of why you chose the particular stocks. Provide time plots of the data and histograms (but try to convey this information in an efficient way by having multiple panels in a given figure). Also, provide tables reporting the summary statistics. If there are large outliers for any of the series, try to determine what happened at the time. Please include this material in your final project too. Make the presentation of graphs as concise as possible (i.e., use multiple panels per page).

4. Portfolio Theory Calculations

a. Using the estimated means and the sample covariance matrix, compute the global minimum variance portfolio for the 10 risky assets. Again, if estimated excess means are negative, use the median estimate or set to zero if excess mean is negative even given median estimate.
b. Using the highest historical mean as a target, compute a second efficient portfolio.
c. Using the two efficient portfolios computed in a. and b., compute the Markowitz bullet (portfolio frontier for 10 risky assets).
d. Using the mean monthly risk-free return, compute the tangency portfolio and the efficient set for the 10 risky assets and a risk-free asset. In addition to the tangency portfolio with short sales, you should also calculate the tangency portfolio with no short sales.
e. Compare weights to the weights for the value-weighted portfolio based on current market capitalization. Discuss why you might expect the weights to be related (especially for the no short-sales case) and why you might expect the weights to be different. (Hint: think about when the market capitalization is measured.)

5. Market Model Calculations

a. Using the monthly return on the market portfolio, estimate the MM for the 10 stocks.
b. For each stock, test the hypothesis that beta=1 versus the alternative that it does not equal 1 using a 5% test. Calculate 95% confidence intervals for the betas.

c. Estimate the beta of the tangency portfolio. Confirm that the beta of a portfolio is the weighted average of the betas of the assets in the portfolio. For the tangency portfolio, test the hypothesis that beta=1 versus the alternative that it does not equal 1 using a 5% test.

6. CAPM Calculations

a. Using the risk-free rate data, calculate excess returns. Run the CAPM regression (MM using excess returns) for each stock.
b. For each stock, test the hypothesis that the intercept term alpha*=0 versus the alternative that it does not. Again, use a 5% test. Also, calculate the 95% confidence intervals for the alphas.
c. Test the CAPM by running a CAPM regression with the excess return on your tangency portfolio as the dependent variable (instead of the return on a given asset, as in part b.). Is alpha* = 0? You should do this test for the tangency portfolio with short sales and then for the tangency portfolio with no short sales.

7. Testing Assumptions

The last thing to do in terms of calculations is test a number of the assumptions made to justify the previous analysis. Always use the 5% level to determine significance.

a. Using the estimated skewness and kurtosis in 3, compute the Jarque-Bera statistic to test normality for each of the 12 return series.
b. Compute the 1,6, and 12 month lag sample autocorrelations and 95% confidence intervals for each return series, excluding the risk-free series. Calculate the Box Q statistic to test the null hypothesis that series are not autocorrelated at up to 12 lags (one statistic for each series). For the two assets with highest contemporaneous correlation with each other, calculate the 1, 6, and 12 month lag sample cross-autocorrelations.
c. For bonus points (i.e., this is an advanced question and is not required), bootstrap the Box Q statistics by drawing from the historical returns with replacement (i.e., imposing the null hypothesis of no correlation). Describe how you set up the bootstrap experiment and how many bootstrap replications you considered. Also, bootstrap the 95% confidence intervals for the autocorrelation considered in part b. Are they different from the standard confidence intervals used in part b. above? Do your inferences about predictability change?
d. Using the residuals from the MM regressions in 5, compute the Jarque-Bera statistic for any of the ten regressions for which the returns were found to be non-Normal in part a. above.
e. Again, using the residuals from the MM regressions in 5, compute the 1,6, and 12 month lag sample autocorrelations and 95% confidence intervals for the residual series for any of the ten regressions for which the returns were found to be serially correlated in part b. above. For these residual series, calculate the Box Q statistic to test the null hypothesis that residuals are not autocorrelated at up to 12 lags (one statistic for each series).
f. Again, using the residuals from the MM regressions in 5, compute the White statistic to test homoskedasticity for each of the ten regressions. Regardless of your findings for the test, construct the heteroskedasticity-consistent standard errors and redo *t*-tests

of beta. Is there a link between your White test results and how much the t-tests change when considering heteroskedasticity-consistent standard errors?

g. Split the sample period in half and test for parameter stability in the CAPM regression parameters using the Chow test. Plot the mean returns from the second half of the sample on the y-axis against the betas from the first half of the sample on the x-axis. Compare to the Securities Market Line (use full sample to get risk free rate and mean return on the market portfolio). How did the CAPM do in prediction? Are deviations from the SML related to results for parameter stability? What if you test the CAPM for a tangency portfolio with weights estimated using the first half of the sample and the test regression using the second half of the sample?

Questions or issues that you should cover in the Final Project write up:

- Which stocks did you expect to have large betas before estimation? Explain in the context of the business cycle behaviour and variance of each stock.
- Were your stocks as representative of the whole as you thought they would be? Think about the role of survivorship bias in interpreting all of your results.
- Do the returns look normally distributed? What do the Jarque-Bera statistics suggest?
- Is there any significant serial correlation for the returns? Do returns appear predictable using past returns?
- Are the means estimated precisely? Are the standard deviations estimated precisely?
- Compare the standard deviation of the risk-free asset to those of the risky assets. Is an assumption of a constant risk-free asset reasonable?
- Are there any negative weights on the 10 stocks in the tangency portfolio? If so, interpret them. How do your results change if you impose the restriction of no short sales? How do these weights correspond to the weights based on market capitalization?
- How does the mean and standard deviation of market portfolio compare to the tangency portfolio and the efficient set given only the 10 risky assets? Is the market portfolio within the efficient frontier or outside of it? What about the efficient set given no short sales? Explain how you constructed the efficient set given no short sales.
- Do the estimated betas for the MM match your priors?
- Is the beta for the tangency portfolio close to 1?
- Are the betas for the CAPM regressions similar to those for the MM regressions?
- Can you reject the hypothesis that any of the alphas are zero? What does this imply for the CAPM? Does it appear that idiosyncratic risk is being priced? How does the CAPM stand up compared to a simple risk/return tradeoff model for each stock? Compare the correlation between the estimated means and betas with the correlation between the estimated means and standard deviations.
- Do the statistical assumptions underlying the MM and CAPM hold?
- Does the CAPM predict expected returns in the second half of the sample?
- Does the SML hold more closely for small or large cap firms?
- Bonus Question: Given confidence intervals for all of the parameters involved, is it possible to statistically reject that any of the stocks are off the SML?
- Based on your results, would you make any recommendations to your reader? Assume your reader is interested in investing.

Format of Final Project:

1. Introduction
    - summarize the main results
2. Data Description
    - where is it from?
    - why did you pick particular stocks?
    - weights based on market cap
    - Time plots of data
    - Histograms
3. CER Model Results
    - List of assumptions of model
    - Table of Means and Standard Deviations for 12 series
    - Table of Jarque-Bera statistics
    - Table of serial correlation results
4. Portfolio Theory Results
    - Describe the optimization problems and the procedure for calculating the Markowitz bullet
    - Plot the Markowitz bullet, the tangency portolio, the efficient set of portfolios given risky assets and a risk-free asset, and the location of the mean and standard deviation of the market portfolio (e.g., S&P500), the equal-weighted portfolio, and the value-weighted portfolio
    - Plot the results under the assumption of no short sales and describe how these calculations were made
    - Present weights based on market cap and compare to tangency portfolio weights.
5. Market Model and CAPM Results
    - MM regression equation and assumptions
    - Table of Betas with confidence intervals for MM
    - CAPM regression equation and assumptions
    - Table of Alphas with confidence intervals for CAPM regressions
    - Table of Jarque-Bera stats for MM residuals
    - Table of serial correlation test results for MM residuals
    - Table of White Test results
    - Table of Chow Test results
    - Plot of CAPM predictions and SML
6. Conclusion
    - Interpret results (how well do assumptions hold, what do the results suggest about diversification, CAPM, etc…)
    - Any advice to the reader?
7. Bibliography
    - Feel free to quote and reference Random Walk Down Wall Street or any other academic source during the write up

Please cover everything mentioned above, but do so using concise and clear writing. You will be graded in part on your ability to present results efficiently.