Some General Information

- This course has an above-average workload. You can assume that there is something due for every class.
- Often, the assignments lead the class in the sense that you work on something first and then we cover it in class. This is a good way to learn, but not everybody likes it.
- Assignments may be done in groups.
- There will be a final exam, and it will be open book, open notes. I am posting last year’s final so you can see what we’ll be covering in this course.
- I consider *Capital Markets and Investments* a co-requisite for this course, but it helps to have taken it before. The key background we need from that course is the CAPM and its connections with regression. If you are taking *Capital Markets* at the same time as this course, you will probably be covering the CAPM right around the time this course starts.

Overview of the Course

This course covers statistical concepts for investment analysis – measuring risk and return and the factors that drive them. The course is organized around three main topics:

- Equity factor models
- Volatility
- Fixed income and credit factors

The first topic will make up about half the course.

We will use specific tools (especially regression), but the emphasis will be on statistical thinking about financial data rather than on technique. In particular, the course will focus on broadly applicable features of market data and the statistical concepts that help us understand them. Each class will pair an idea from investment analysis with one or more statistical tools. All of the concepts and tools developed in the course will be closely aligned with industry practice, and we will frequently draw on both industry and academic research. No familiarity with statistics beyond the MBA core course is required.
Prerequisites

The only formal prerequisite is the core class in Managerial Statistics. If you exempted out of the core statistics class, be sure you are comfortable interpreting regression output. I will assume familiarity with financial markets and terminology at the level of Capital Markets and Investments, so I strongly recommend taking that class in parallel with this one if you have not taken it before. This course will also have points of contact with the core courses in Corporate Finance and Business Analytics.

Course Work and Grading

There is no textbook for the course. The course will be taught from lecture notes and background readings.

You should assume that there is an assignment due for every class. Homework assignments may be done individually or in groups of up to three people, and you may work with different teams on different assignments. If you would like help finding a team, email the TA. (We will discuss this on the first day of class.)

Assignments will be a mix of data analysis, thought questions, and reading. If you lose points on a data analysis problem, you have the option redo your analysis and resubmit your assignment for full credit.

In calculating your course grade, I will drop your lowest homework score.

We will have a final exam, which will be based closely on the homework assignments and class material. My intention is that if you have kept up in class and understood the homework assignments, you will find the final exam straightforward; and if you have not kept up in class or not been conscientious about the homework, you will find the final exam difficult.

Class participation – meaning “present, prepared, participating” – is important for learning and will help keep the class interesting for all of us.

Grades will be based on the following weights:

- Assignments – 60%
- Final exam – 30%
- Class participation – 10%

Getting Help Outside of Class

I am available. I will hold regular office hours (barring unexpected constraints), but you should feel free to email me or talk to me after class to set up a different time. We will also have a TA for the course, and the TA’s office hours and contact information will be posted on Canvas.

Software
All data analysis assignments are designed to be done in Excel, so there is no requirement to use statistical software. However, if you happen to be familiar with a statistical package or want to experiment with other software, you are free to use any tools you wish. Many products offer free trial versions.

Course Outline

This is an overview of the topics in the course. Consult the course page on Canvas for detailed information on assignments.

1. Introduction
   - Course overview
   - Quantitative investment management
   - Review of the CAPM and regression – what works and what doesn’t work in the CAPM and why it’s relevant to investment analysis
   - Decomposing risk in a single-factor model

2. Equity factor models
   - Why factor models?
   - Fama-French factors – size and value
   - Alternative beta versus alpha

3. Equity factor models and performance evaluation
   - Momentum
   - Performance measures – Sharpe ratio and information ratio
   - Evaluating fund performance through regression
   - Other equity factors

4. Risk models
   - Factor models of risk
   - Partial correlation
   - Quantitative, fundamental, macroeconomic, and statistical factors
   - Risk factors versus alpha factors

5. Factor reduction through principal components analysis
Why a small number of factors often does the job
Calculating and interpreting principal components

6. Long-term trends
- Long-horizon predictive regressions using P/E ratios or dividend yields
- Trend following and momentum
- Applications to sector allocation and global macro strategies

7. Properties of time series data
- Trend, seasonality, stationarity
- Autocorrelation and autoregressive models
- Cointegration versus correlation and application to pairs trading

8. Volatility I
- Measuring volatility: realized, implied, VIX
- Persistence; GARCH
- Leverage effect

9. Volatility II
- Volatility risk premium; covered calls
- Low-risk anomaly
- Risk parity

10. Dynamics of interest rates
- Properties of interest rate data
- Yield curve risk
- Nelson-Siegel model; level, slope, and curvature
- Forecasting the coefficients
- Incorporating macro factors

11. Credit risk and credit scoring I
- Altman Z-score; discriminant analysis

12. Credit risk and credit scoring II
- Logistic regression
- Credit ratings

The final exam will be held during exam period.
Final Exam
(This is last year's final exam, for your reference.)

- Write all answers directly on the exam
- Do not remove the staple that holds the pages together
- The exam lasts 3 hours. Do not discuss the exam until the 3 hours are up, even if you finish early
1. Consider the regression
\[ R_i - R_F = \alpha + \beta (R_m - R_F) + \epsilon \]
in which \( R_i \) is the return on an individual stock, \( R_F \) is the risk-free interest rate, and \( R_m \) is the return on the market. The key assertion of the CAPM is which of the following:

(a) The residual \( \epsilon \) is uncorrelated with the market return \( R_m \)
(b) The residual \( \epsilon \) is uncorrelated with the excess return \( R_m - R_F \)
(c) The intercept \( \alpha \) is equal to zero
(d) A higher beta implies that excess returns on the stock have greater sensitivity to excess market returns

2. Consider again the regression in question 1. Suppose you have the output of this regression. Which of the following is the most relevant to the level of stock-specific risk for the stock in the regression?

(a) 1-\( R^2 \)
(b) 1- \( \beta \)
(c) |\( \alpha \)|
(d) \( \epsilon^2 \)

3. A regression of the monthly returns of the PG20 fund produces the following results:

<table>
<thead>
<tr>
<th>Regression Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multiple R 0.97</td>
</tr>
<tr>
<td>R Square 0.95</td>
</tr>
<tr>
<td>Adj R Square 0.95</td>
</tr>
<tr>
<td>Standard Error 1.03</td>
</tr>
<tr>
<td>Observations 417</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ANOVA</th>
</tr>
</thead>
<tbody>
<tr>
<td>df</td>
</tr>
<tr>
<td>Regression 5</td>
</tr>
<tr>
<td>Residual 411</td>
</tr>
<tr>
<td>Total 416</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Coefficients</th>
<th>Std Err</th>
<th>t Stat</th>
<th>P-value</th>
<th>Lower 95%</th>
<th>Upper 95%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept 0.32</td>
<td>0.10</td>
<td>3.35</td>
<td>0.00</td>
<td>0.43</td>
<td>0.81</td>
</tr>
<tr>
<td>rf 0.93</td>
<td>0.18</td>
<td>5.09</td>
<td>0.00</td>
<td>0.57</td>
<td>1.29</td>
</tr>
<tr>
<td>mktrf 0.87</td>
<td>0.01</td>
<td>72.06</td>
<td>0.00</td>
<td>0.84</td>
<td>0.89</td>
</tr>
<tr>
<td>smb 0.59</td>
<td>0.02</td>
<td>34.18</td>
<td>0.00</td>
<td>0.56</td>
<td>0.62</td>
</tr>
<tr>
<td>hml 0.38</td>
<td>0.02</td>
<td>20.63</td>
<td>0.00</td>
<td>0.35</td>
<td>0.42</td>
</tr>
<tr>
<td>umd 0.16</td>
<td>0.01</td>
<td>14.19</td>
<td>0.00</td>
<td>0.14</td>
<td>0.18</td>
</tr>
</tbody>
</table>
The underlying data for the fund’s return and for the factors is in percent. For example, the first row of data looks like this

<table>
<thead>
<tr>
<th>rf</th>
<th>mktrf</th>
<th>smb</th>
<th>hml</th>
<th>umd</th>
<th>PG20</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.36</td>
<td>-3.99</td>
<td>4.79</td>
<td>4.25</td>
<td>3.94</td>
<td>1.96</td>
</tr>
</tbody>
</table>

meaning that in the first month, the fund returned 1.96% while the market’s excess return was -3.99%. Some additional information about the underlying monthly factor returns is provided in the tables below:

After accounting for the factors in the regression, what is PG20’s alpha, measured as a monthly return?
4. This question uses the data in question 3. What is the average monthly return of the PG20 fund for the period covered by the regression?

5. This question uses the data in question 3. Suppose we use the regression coefficients and the factors in the regression to define a benchmark for PG20. What is the fund’s information ratio relative to this benchmark?

6. Which of the following statements is best supported by the data in question 3?

(a) Relative to the overall market, the PG20 fund is more invested in small stocks.
(b) More of the fund’s average return comes from momentum than from value.
(c) The fund is market-neutral.
(d) Most of the fund’s risk (as measured by the standard deviation of monthly returns) is not captured by the factors in the regression.
7. For this question use the following information:

<table>
<thead>
<tr>
<th>Descriptive statistics of monthly returns</th>
<th>Regressions on SP500 returns</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>beta</td>
</tr>
<tr>
<td>SP500 CAT CMG CSX DPS</td>
<td>CAT</td>
</tr>
<tr>
<td>Median 1.5% 0.6% 3.4% 0.7% 1.3%</td>
<td>0.27</td>
</tr>
<tr>
<td>Stdev 3.8% 9.4% 10.0% 6.9% 4.2%</td>
<td>0.27</td>
</tr>
<tr>
<td>Skewness -0.01 0.78 -0.69 0.02 -0.20</td>
<td>0.44</td>
</tr>
<tr>
<td>Kurtosis 3.66 4.74 4.16 3.29 3.15</td>
<td>0.19</td>
</tr>
</tbody>
</table>

The table in the upper left shows descriptive statistics on four stocks and the S&P500. The table in the lower left shows their correlations. The table in the upper right shows the results of single-variable regressions of individual stock returns on S&P500 returns. The italicized numbers under the beta and intercept are their standard errors. Each s_e is the standard deviation of the residuals for the regression.

What is the correlation between the residuals of CAT and CSX?
8. This question uses data on the Case-Shiller house price indices for 19 U.S. cities. The data is monthly from January 2000 to July 2013. The following two charts show the first two principal components for the returns across the 19 cities:

Based on these charts, which city do you think is most correlated with Phoenix? Explain how you reach your conclusion.
9. Consider the following chart from a recent research report:

Exhibit 8

Co-integrated Earnings Are More Common Among Sectors

<table>
<thead>
<tr>
<th>Results of Cointegration Tests for Aggregate Sector Earnings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Null Hypothesis: these Series Are NOT Cointegrated</td>
</tr>
<tr>
<td>The Table Below Shows the Pairwise Rejection Significance Levels</td>
</tr>
<tr>
<td>Critical Values for These Tests Are Non-Standard</td>
</tr>
<tr>
<td>July 1967 Through July 2006</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Energy</th>
<th>Materials</th>
<th>Industrials</th>
<th>Discretionary</th>
<th>Staples</th>
<th>Health Care</th>
<th>Financials</th>
<th>Technology</th>
<th>Telecoms</th>
<th>Utilities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Energy</td>
<td>10% level</td>
<td>10% level</td>
<td>5% level</td>
<td>10% level</td>
<td>10% level</td>
<td>na</td>
<td>na</td>
<td>10% level</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Materials</td>
<td>5% level</td>
<td>5% level</td>
<td>5% level</td>
<td>5% level</td>
<td>5% level</td>
<td>na</td>
<td>na</td>
<td>5% level</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Industrials</td>
<td>5% level</td>
<td>5% level</td>
<td>5% level</td>
<td>5% level</td>
<td>5% level</td>
<td>na</td>
<td>na</td>
<td>1% level</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Discretionary</td>
<td>1% level</td>
<td>1% level</td>
<td>1% level</td>
<td>na</td>
<td>na</td>
<td>na</td>
<td>na</td>
<td>1% level</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Staples</td>
<td>1% level</td>
<td>5% level</td>
<td>na</td>
<td>na</td>
<td>na</td>
<td>insignif</td>
<td>na</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Health Care</td>
<td>1% level</td>
<td>5% level</td>
<td>na</td>
<td>na</td>
<td>na</td>
<td>insignif</td>
<td>na</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Financials</td>
<td>na</td>
<td>na</td>
<td>5% level</td>
<td>na</td>
<td>na</td>
<td></td>
<td>na</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Technology</td>
<td>na</td>
<td>na</td>
<td></td>
<td>na</td>
<td>na</td>
<td></td>
<td>na</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Telecoms</td>
<td>na</td>
<td>na</td>
<td></td>
<td>na</td>
<td>na</td>
<td></td>
<td>na</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Utilities</td>
<td>na</td>
<td>na</td>
<td></td>
<td>na</td>
<td>na</td>
<td></td>
<td>na</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Suppose earnings for the Health Care sector have recently outpaced earnings for other sectors. Which of the following statements is best supported by the chart?

(a) Energy earnings are likely to grow
(b) Staples earnings are likely to grow
(c) Energy earnings are likely to decline
(d) Staples earnings are likely to decline
(e) Health Care earnings are likely to continue to grow
10. Suppose you are calculating an exponentially weighted moving average with decay parameter $\lambda$. What is the advantage of using a larger value of $\lambda$ and what is the advantage of using a smaller value?

11. This question refers to Appendix 3. Explain the pattern in the main chart (the GARCH values) and explain how it relates to the other two charts. Be succinct and precise.
12. Consider the following GARCH(1,1) model based on daily data.

\[
\sigma_t^2 = (1 - \alpha - \beta)\sigma^2 + \alpha R_{t-1}^2 + \beta \sigma_{t-1}^2
\]

Suppose you estimate \( \alpha = 0.07 \) and \( \beta = 0.90 \). How long would you forecast it will take for the current level of variance to get halfway to its long-run level?
13. Recall that in the Nelson-Siegel model of the yield curve, the Level factor is the constant 1 and the Slope and Curvature factors are defined by

\[
\text{Slope}(t) = \frac{1 - \exp(-\lambda t)}{\lambda t} \quad \text{Curv}(t) = \frac{1 - \exp(-\lambda t)}{\lambda t} - \exp(-\lambda t),
\]

with time \( t \) measured in months. (This expression for Curv corrects a typo in the slides.) For this problem, \( \lambda = 0.0609 \). For the coefficients on these three factors, we have fit the following AR(1) models based on monthly observations:

\[
\begin{align*}
\beta_{1,t+1} &= 0.208 + 0.968 \beta_{1,t} + \epsilon_{t+1} \\
&\quad (0.113) \quad (0.015) \\
\beta_{2,t+1} &= -0.012 + 0.984 \beta_{2,t} + \epsilon_{t+1} \\
&\quad (0.041) \quad (0.015) \\
\beta_{3,t+1} &= -0.003 + 0.892 \beta_{3,t} + \epsilon_{t+1} \\
&\quad (0.061) \quad (0.033)
\end{align*}
\]

The numbers in parentheses indicate standard errors. The residual terms are assumed uncorrelated across the three models. Suppose the current levels for the three betas are 5.39, 0.88, and -0.73. What is your forecast of where the 2-year yield will be in one month?
14. This question refers to Exhibit 4 on the Credit Suisse page (Appendix 1) with the heading “Principal component analysis of the JGB markets.” (JGB = Japanese government bond.) Compare the values of PC1 and PC2 in 2012 with their values in 2005. Write a sentence interpreting the change (without using the term “principal component”).

15. A logistic regression using the Altman bankruptcy data produces the following output:

<table>
<thead>
<tr>
<th>Variables</th>
<th>Constant</th>
<th>X1</th>
<th>X4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coefficients</td>
<td>4.5320</td>
<td>-0.0881</td>
<td>-0.0238</td>
</tr>
<tr>
<td>Std error</td>
<td>1.3326</td>
<td>0.0290</td>
<td>0.0082</td>
</tr>
<tr>
<td>p-value</td>
<td>0.0007</td>
<td>0.0024</td>
<td>0.0037</td>
</tr>
<tr>
<td>Log-likelihood</td>
<td>-16.23</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number valid obs</td>
<td>66</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total obs</td>
<td>66</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Recall that the variables have the following definitions:

- \( X_1 = \frac{\text{Working capital}}{\text{Total assets}} \)
- \( X_2 = \frac{\text{Retained earnings}}{\text{Total assets}} \)
- \( X_3 = \frac{\text{Earnings before interest and taxes}}{\text{Total assets}} \)
- \( X_4 = \frac{\text{Market value equity}}{\text{Book value of total liabilities}} \)
- \( X_5 = \frac{\text{Sales}}{\text{Total assets}} \)

Consider a firm with the following variable values:

<table>
<thead>
<tr>
<th>X1</th>
<th>X2</th>
<th>X3</th>
<th>X4</th>
<th>X5</th>
</tr>
</thead>
<tbody>
<tr>
<td>24</td>
<td>3.3</td>
<td>-3.5</td>
<td>20.9</td>
<td>1.1</td>
</tr>
</tbody>
</table>

What is the firm’s one-year probability of bankruptcy, according to the logistic regression model?
16. This question uses the data from the previous question. What would be the impact on the firm’s bankruptcy probability if its stock price dropped by 10%, with everything else held fixed?

17. This question refers to Appendix 2. Write a few sentences explaining the statement in the lower left that “Quality stocks are increasingly growth biased, while their beta deficit to junk stocks has narrowed (Exhibit 7).” You should explain the steps one would go through to arrive at this conclusion and what it means.
18. The Barra US Equity risk model includes a factor called Non-Linear Size, which is their Size factor cubed. Barra’s documentation includes the following statement:

The Non-Linear Size (NLS) factor captures non-linearities in the payoff to the Size factor across the market-cap spectrum. This factor is based on a single raw descriptor: the cube of the Size exposure. However, since this raw descriptor is highly collinear with the Size factor, it is orthogonalized with respect to Size. This procedure does not affect the fit of the model, but does mitigate the confounding effects of collinearity, while preserving an intuitive meaning for the Size factor.

Given two correlated factors (in this case, Size and NLS) how could you replace them with two other factors to remove collinearity without affecting the fit of the model?

19. Which time series do you think is more highly persistent: the squared values of daily market returns or the cubed values of daily market returns? Justify your answer.
Principal component analysis in the JGB markets

Meanwhile, PCA also shows, as did our examination above, that sensitivity to changes in yield level in the JGB markets is tending to increase. Exhibits 3 and 4 below show the contribution ratio indicated when using PCA to extract the PC factors from daily moves in the JGB curve. The Eigenvectors of PC1 through PC3, consistent with many empirical studies, indicate that when separating changes in the yield curve into level changes (PC1), slope changes (PC2), and curvature changes (PC3), in our view (Exhibit 5).

Exhibit 3: Trends in PC1 through PC3 contribution rate (observation period: 20 business days)

Exhibit 4: Trends in PC1 through PC3 contribution rate (observation period: 60 business days)

Source: Credit Suisse

Exhibit 5: Eigenvectors of PC1 through PC3 in the JGB markets (average value in 60 business days)

Source: Credit Suisse

Such an interpretation of these PC factors becomes obvious when comparing against actual market moves. As shown in Exhibit 6 below, the PC1 factor score correlates very highly with the 10-year JGB yield. This is consistent with the interpretation of PC1 as an expression of changes in yield level (i.e., level factor). The same can be shown for PC2 and PC3 (Exhibits 7 and 8). In other words, the PC2 factor score correlates highly with the slope of the yield curve (the 2yr/10yr spread) and the PC3 factor score correlates highly with the curvature of the yield curve (the 2yr/10yr/30yr butterfly spread). As shown in Exhibits 3 and 4 above, those three PC factors combined explain at least 80% of yield curve movement.
**Appendix 2**

**Morgan Stanley**

**Equity Quality Performance**

Equity quality can be a subtle characteristic to define. We constructed an equity quality model that splits stocks into four quality cohorts: Quality, Moderate Quality, Low Quality and Junk. This model uses long-run data (e.g., stability of earnings per share and dividends) when it is available for stocks, but for newer stocks it substitutes more immediate information (e.g., share base turnover and equity beta); see US Equity Strategy: The Quality / Junk Debate, January 17, 2011 for more details on the model. In particular, we do not use past performance or back-tested future performance of stocks to make quality classifications. While the quartile classifications may potentially change each month, turnover among quartiles is less than 6% per month, with less than 1% of stocks moving two or more quartiles. As with equity style, we compute quality quartile performance based on equal-weight returns of stocks in the quartile each month.

By construction, the quality quartiles contain differing degrees of equity market beta and size sensitivity. We therefore compute quality residuals net of equity market, size and style. Given the uniqueness of our quality model, we felt it was important to use publicly available risk factor returns (rather than proprietary metrics) to compute residuals, so the focus remains on our model and not on the details of the hedging.

**Quality stocks are increasingly growth biased, while their beta deficit to junk stocks has narrowed** (Exhibit 7). While some investors may fuse quality and value, they are distinctive notions in our models. The evolution of the quality-junk spread from a value bias to a growth bias is not connected to AAPL’s ascension to high quality after initiating a dividend, as our quality quartile returns are computed equal-weighted, rather than cap-weighted.

**Exhibit 7**

The Growth Bias of Quality Stocks Continues to Increase, while their Beta Differential Narrows

Systematic Factor Exposures of MS Quality-Junk Spread
Uses the Fama-French Three-Factor Model

Source: Factset, Morgan Stanley Research

Quality stocks have outperformed junk stocks by 314bp over the last 12 months (Exhibit 8). While quality-junk spreads fluctuate from month to month, quality has generally outperformed junk during the Crisis recovery. Low interest rates engendered by the Fed’s QE program have generally favored quality stocks (which often pay stable and/or growing dividends according to our model). Our view remains that absent a synchronous global economic expansion, we are unlikely to have a sustained rally by the 3rd and 4th quartiles of our quality model, as typically this is pressaged by high bankruptcy risk among a substantial group of equities, which is not the case today.

**Exhibit 8**

High Quality Stocks Outperformed Junk by 314bp in the Last 12 Months on a Risk-Adjusted Basis
Appendix 3

S&P 500 COMPOSITE GARCH VOLATILITY GRAPH

Volatility Prediction for Monday, December 2: 9.36% (-0.23)

Volatility Summary Table

<table>
<thead>
<tr>
<th></th>
<th>Avg Week Vol:</th>
<th>Avg Month Vol:</th>
<th>1 Week Pred:</th>
<th>1 Month Pred:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Closing Price (USD):</td>
<td>51,805.81</td>
<td>10.74%</td>
<td>10.15%</td>
<td></td>
</tr>
<tr>
<td>Avg Vol:</td>
<td>9.61%</td>
<td>10.74%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Min Vol:</td>
<td>7.42%</td>
<td>7.88%</td>
<td>12.87%</td>
<td></td>
</tr>
<tr>
<td>Max Vol:</td>
<td>17.55%</td>
<td>22.39%</td>
<td>14.45%</td>
<td></td>
</tr>
</tbody>
</table>