The Capital Asset Pricing Model
Theory, Econometrics, and Evidence

S. 2011
Department of Business Studies
Aarhus School of Business
Aarhus University
Abstract

In this thesis three aspects of the CAPM model are investigated. The first aspect is the theoretical background of the model. Here, mean-variance analysis (MVA) is thoroughly examined. We first present mathematical arguments from utility theory that can motivate the implementation of MVA. Then, we examine efficient portfolios in a mean-standard deviation space assuming there is no risk-free asset. We show the incentive to diversify ones portfolio and derive the efficient frontier consisting of the portfolios with the maximum expected return for a given variance. Using mathematical and economic arguments, we find out that the market portfolio consisting of all risky assets is mean-variance efficient. We then include a riskless asset in the analysis and get the Capital Asset Market Line (CML) in a mean-standard deviation space. We argue that this is the efficient frontier when a risk-free rate exists. We also present the separation theorem which implies that all investors will maximize utility in some combination between the risk-free asset and the market portfolio. Based on the CML, we derive the Capital Asset Pricing Model (CAPM) in three different ways. The first two represent the original approaches from the architects behind the model. The last approach extends the formal derivation of the efficient frontier when only risky assets exists to include the risk-free rate. We find that the CAPM relates the expected return on any asset to its beta. We argue that when investors only care about expected return and variance, beta makes sense as a risk measure. As beta is based on the covariance of returns between an asset and the market portfolio, it follows that CAPM only rewards investors for their portfolios responsiveness to swings in the overall economic activity. We find that this makes sense, as rational investors can diversify away all but the systematic risk of their portfolios.

In the second part of the thesis, the econometric methods for testing the CAPM are developed. First, the traditional model is rewritten in order to work with excess returns. We then focus on testing the mean-variance efficiency of the market portfolio. We impose the assumption that returns are independent and identically distributed and jointly multivariate normal. Based on this assumption, we derive the joint probability density function (pdf) of excess returns conditional on the market risk premium. Using this pdf we first derive maximum likelihood estimators of the market model parameters. We then show that they are, in fact, equivalent to the ordinary least squares estimators. A number of different test statistics are derived based on these estimators. The first is an asymptotic Wald type test. We then transform this test into an exact $F$ test. Moreover, we develop an asymptotic likelihood ratio test including a corrected version with better finite-sample properties. Also, noting that the above distributional assumptions are rather strict, we use the Generalized Method of Moments (GMM) framework to develop a test robust to heteroskedasticity, temporal dependence and non-normality. Finally, we present some cross-sectional tests of other implications of the CAPM. Specifically, we develop $t$ statistics to check whether the empirical market risk premium is significant and positive and whether other risk measures than beta have explanatory power regarding expected excess returns.

The third part of the thesis is an empirical study. It starts out by discussing a number of relevant topics regarding the implementation of the statistical tests. In specific, we discuss the choice of proxies, the sample period length and frequency, and the construction of the dependant variable. Then, the tests are carried out on a 30 year sample of American stocks. For the overall period, we cannot reject the mean-variance efficiency of the proxy for the market portfolio. However, for the subperiods of 5 years, the results are not so clear-cut. We also find that the empirical risk premium is not significant. The last point clearly contradicts the CAPM framework.
# Table of Contents

**CHAPTER 1 INTRODUCTION** .......................................................................................................................................................................................... 1
  1.1 PURPOSE AND RESEARCH QUESTIONS .......................................................................................................................... 1
  1.2 DELIMITATIONS ................................................................................................................................................................. 2
  1.3 NOTATION AND SOURCES ................................................................................................................................................ 2
  1.4 STRUCTURE ........................................................................................................................................................................ 2

**CHAPTER 2 THEORY AND DERIVATIONS** ......................................................................................................................................................... 4
  2.1 MEAN-VARIANCE ANALYSIS ................................................................................................................................................ 4
  2.2 THE MINIMUM VARIANCE FRONTIER .................................................................................................................................. 7
  2.3 THE CAPITAL MARKET LINE .............................................................................................................................................. 14
  2.4 DERIVATIONS OF THE CAPM ........................................................................................................................................... 17
    2.4.1 Sharpe’s derivation ................................................................................................................................................... 17
    2.4.2 Lintner’s derivation .............................................................................................................................................. 19
    2.4.3 Derivation by solving a minimization problem ........................................................................................................ 22
  2.5 THE SECURITY MARKET LINE ........................................................................................................................................ 25
    2.5.1 Beta as a risk measure ........................................................................................................................................ 26

**CHAPTER 3 ECONOMETRIC TECHNIQUES FOR TESTING THE CAPM** ................................................................................................................. 28
  3.1 THE MARKET MODEL ...................................................................................................................................................... 28
  3.2 THE STANDARD TESTS ................................................................................................................................................ 36
  3.3 SIZE OF THE TESTS ....................................................................................................................................................... 42
  3.4 POWER OF THE TESTS ................................................................................................................................................ 44
  3.5 THE ASSUMPTIONS OF IID AND JOINTLY NORMAL RETURNS ...................................................................................... 46
  3.6 CROSS-SECTIONAL REGRESSIONS .................................................................................................................................... 50

**CHAPTER 4 EMPIRICAL STUDY** ............................................................................................................................................................................. 52
  4.1 LITERATURE REVIEW .................................................................................................................................................. 52
  4.2 ISSUES IN IMPLEMENTING THE TESTS ............................................................................................................................... 53
    4.2.1 The choice of proxies ............................................................................................................................................. 53
    4.2.2 Period length and sampling frequency .................................................................................................................. 54
    4.2.3 Portfolio construction ............................................................................................................................................ 54
  4.3 DATA SELECTION .......................................................................................................................................................... 54
  4.4 EMPIRICAL TESTS ....................................................................................................................................................... 55
    4.4.1 PARAMETER ESTIMATES OF THE MARKET MODEL .............................................................................................. 55
    4.4.2 TIME-SERIES TESTS OF THE INTERCEPT ........................................................................................................... 57
    4.4.3 CROSS-SECTIONAL TESTS ........................................................................................................................................ 60
    4.4.4 GRAPHICAL INTERPRETATION AND CONCLUSION ............................................................................................. 61

**CHAPTER 5 ALTERNATIVE ASSET PRICING MODELS** ................................................................................................................................. 65
  5.1 THE BLACK VERSION .................................................................................................................................................. 65
  5.2 ARBITRAGE PRICING THEORY ........................................................................................................................................ 65

**CHAPTER 6 CONCLUSION** ................................................................................................................................................................. 67

**BIBLIOGRAPHY** ................................................................................................................................................................................. 69

**APPENDIX** .................................................................................................................................................................................. 73
  APPENDIX A.1 ........................................................................................................................................................................ 73
  APPENDIX A.2 ........................................................................................................................................................................ 74
  APPENDIX A.3 ........................................................................................................................................................................ 75
  APPENDIX B.1 ........................................................................................................................................................................ 76
List of Figures

Figure 2.1 Risk-return frontiers in the two-asset case .............................................................................................................8
Figure 2.2 a: Risk-return frontiers in the three-asset case ......................................................................................................8
Figure 2.2 b: Minimum variance frontier with N assets ..........................................................................................................8
Figure 2.3 The optimal portfolio choice in a mean-standard deviation space ......................................................................15
Figure 2.4 Combining the market portfolio with some random risk asset ...........................................................................18
Figure 2.5 The Security Market Line .....................................................................................................................................26
Figure 3.1: Complementary CDFs of as chi-square distribution (left) and a central F distribution (right) .............................42
Figure 4.1: Annualized excess returns vs. beta in the 30 year period and the 10 year subperiods using the CRSP value-weighted proxy of the market portfolio .................................................................63
Figure 4.2: Annualized excess returns vs. beta in the 5 year subperiods using the CRSP value-weighted proxy of the market portfolio ........................................................................................................................................64
List of Tables

Table 3.1: True size of the standard asymptotic tests with an asymptotic size of 5 %....................................................43
Table 3.2: Power of the exact F-test with a size of 5 % given different properties of the tangency portfolio. ..................45
Tabel 4.1: Parameter estimates from the market model of the 10 size-sorted portfolios in the 30 year period and 10 year subperiods. ...............................................................................................................................................................56
Tabel 4.2: Parameter estimates from the market model of the 10 size-sorted portfolios in the 5 year subperiods. ......57
Tabel 4.4: Tests statistic values and p-values for the tests of the null-hypothesis α = 0 using the CRSP value-weight proxy of the market portfolio.................................................................58
Tabel 4.5: Tests statistic values and p-values for the tests of null-hypothesis α = 0 using the CRSP equal-weight proxy for the market portfolio. ...........................................................................................................59
Table 4.6: Tests statistic values and p-values for the tests of the cross-sectional model using the CRSP value-weight proxy for the market portfolio. .................................................................60
Tabel 4.7: Tests statistic values and p-values for the tests of the cross-sectional model using the CRSP equal-weight proxy for the market portfolio. .....................................................................................61
Chapter 1 Introduction

In 1952 Harry Markowitz made a historical contribution to financial mathematics with his classic article “Portfolio Selection”. In the article, he incorporated the quantification of risk in the portfolio choice problem. He developed a framework where investors who like wealth and dislike risk would hold mean-variance efficient portfolios\(^1\). Building on his work, Sharpe (1964) and Lintner (1965 b) almost simultaneously developed a model to price capital assets. Sharpe first agreed that Lintners findings superseded his. Later, Fama clarified that their models were, in fact, equivalent (Fama, 1968). The equation they derived has later been christened the Capital Asset Pricing Model\(^2\) (CAPM). This model relates expected return to a measure of risk that incorporates what some consider the “only free lunch of economics”: diversification. This measure, now known as beta, used the theoretical result that diversification allows investors to escape the company specific risk. Thus, they should only be rewarded for their portfolio’s sensitivity to the level of economic activity. Since the original papers, a voluminous literature on the topic has accumulated. A lot of it has dealt with testing the model’s performance on financial data. Even though it is no secret that CAPM’s empirical record is poor, it is still widely used in both academics and in real-world applications (Fama, 2004). In fact, surveys have shown that it is the most popular model among practitioners for estimating expected return (Bartholdy and Peare, 2003). Obviously, the major attraction of the CAPM is its economic intuition and ease of use. Nonetheless, the lethal blows researchers such as Fama and French (1992, 1996) have dealt to CAPM suggest the problems inherent in applying it.

1.1 Purpose and Research Questions

Given the popularity of the CAPM, this thesis seeks to provide the tools needed to empirically investigate and fully capture the model. The exposition will be crowned with updated evidence, to test whether more recent data justifies the model’s wide use. To fulfill these ambitions, a thorough account of the model’s theoretical anchor will be given. The thesis will combine mathematical rigor with economic intuition, and it will arrive at a number of theoretical results. These results will add to the understanding of the central elements of the model and be useful when conducting empirical analysis. After the theoretical background has been established, the econometric methods for testing the CAPM are developed. We will focus on time-series regressions in a multivariate setting, but a cross-sectional model is also investigated. With the statistical techniques settled, we can investigate the empirical performance of the model. We finalize with a short introduction to alternative models of asset pricing. In short, the thesis seeks answers to the following research questions:

1. What is the theoretical foundation of the Capital Asset Pricing Model?
2. How can econometric techniques be applied to test the implications of the Capital Asset Pricing Model?
3. Does the Capital Asset Pricing Model succeed in describing financial return data?

---

\(^1\) A formal introduction to this concept is given in chapter 2.
\(^2\) The 1964 paper has doubtlessly contributed to Sharpe receiving the nobel price in 1990.
1.2 Delimitations
One important version of the CAPM, other than the Sharpe-Lintner model, is the one developed by Black (1972). He relaxes the assumption regarding the existence of a riskless asset and instead allows for unlimited short selling. Though this model has been influential on the field of asset pricing, we will refrain from diving into it. We will instead focus on the traditional version. However, given its importance, a short account will be given in chapter 5. Furthermore, we will not dwell on the applications of the CAPM. The practical implementations of the model are already well established. Nevertheless, we will give examples of its use when appropriate. Finally, in the empirical study, we do not test the explanatory power of other risk measures beyond beta, even though such effects have been documented. For references see chapter 4.

1.3 Notation and Sources
First, \( N \) will generally denote the number of assets or portfolios under consideration. Second, \( E[R_1] \) and \( \sigma_1^2 \) will denote the expected, total return and variance on asset or portfolio 1 respectively. Further definitions will be provided when appropriate. It should also be mentioned that the notation regarding covariances can seem a bit inconsistent. In chapter 2 \( \sigma_{XY} \) is mostly used, where \( Cov(X,Y) \) is applied in chapter 3. The choices reflect the most convenient way of expressing the mathematical relation at hand. In the mathematical derivations bold characters indicate vectors or matrices. Also, an apostrophe ‘ on a matrix or vector denotes that it is transposed. Suppose for example that \( H \) is a \( N \times 1 \)-vector. Then \( H' \) is its transpose with the dimensions \( 1 \times N \). Finally, note that expressions such as \( f'(x) \) and \( \partial g/\partial x \) will be used to denote the derivative of \( f(x) \) and \( g \) with respect to \( x \) respectively.

The theoretical results and derivations will be based on academic articles and textbooks. A source will always be placed immediately after the section using it. Thus, the text and formulae in each section is always based upon the next source listed.

1.4 Structure
As indicated, the thesis is divided into three main parts: theory, econometrics, and evidence. They constitute the second, third, and fourth chapter. The last two chapters deal with alternative models and the conclusions of the thesis respectively.

- Chapter 2 – A rigorous account of the model’s theoretical background is given. This involves an introduction to mean-variance analysis and its extensions in efficient frontier theory. Furthermore, to mirror its origins, the CAPM will be derived using different approaches. This also allows for a more careful interpretation of its implications.

- Chapter 3 – The statistical framework for testing the model is developed. First we work with time-series regression. Maximum Likelihood estimators and their distributions are derived, and test statistics are formed. Also, we develop a test in the Generalized Method of Moment framework that is robust to non-normality, temporal dependence, and heteroskedasticity. Finally cross-sectional regressions are considered.

- Chapter 4 – The tests developed in chapter 3 are applied on a sample of American securities. Relevant discussions regarding data selection and implementation of the tests will be provided. Also, a careful interpretation of the results is given including a graphical analysis.
• Chapter 5 – Alternative asset pricing models are briefly examined. We consider Black’s version of the CAPM, and a multifactor pricing model is also presented.

• Chapter 6 – The conclusions regarding the research questions are summarized into an overview of the central findings of the thesis.
Chapter 2 Theory and Derivations

This chapter will provide a rigorous exposition of the theoretical anchor of the CAPM. First, the foundation is laid by introducing mean-variance analysis. Then the efficient frontier is derived with and without a riskless asset. With these tools in place, we can derive the CAPM in three different ways each contributing to a more thorough understanding of the model. When the equation has been derived, we finally turn to the security market line, and a more intuitive interpretation of the CAPM is given. First, we need to lay some ground rules by listing the rather strict assumptions of the CAPM framework.

In their classic papers Sharpe (1964) and Lintner (1965 b) made two key assumptions for developing the model:

1) There exists a risk-free asset that investors may borrow or lend any amount of at the same rate.

2) Investors assign the same probability distributions to the end of period values of a given asset. This implies that investors have homogenous expectations regarding means, variances and covariances.

They also assumed that investors are wealth loving risk averters, and that the market is purely competitive, frictionless and without taxes. Obviously a constant risk-free rate is not realistic. First, it is highly doubtful that any investment is entirely risk-free. Of course, approximations as government bonds and deposit accounts exist. However, the interest rate does normally vary between lending and borrowing, and it does normally depend on the borrower/lender. Furthermore, investors are humans with subjective assessments of the likelihood of events. Thus, the assumption of homogenous expectations also seems unreasonable. The assumptions regarding risk aversion and the market are probably more realistic. The notion of frictionless implies that there are no transaction costs, and a purely competitive market means that individual investors cannot affect prices by their actions. Furthermore, if incomes taxes and capital gain taxes would be the same, the major implications of the model still hold. The technical assumptions that all assets are marketable and infinitely divisible should also be mentioned (Elton and Gruber, 1995).

2.1 Mean-Variance Analysis

Mean-Variance Analysis (MVA) is the process of selecting a portfolio that provides maximum expected return for a given variance and minimum variance for a given expected return. It is based on the assumption that investors like wealth and dislike risk (i.e. are risk averse). To get a sense of its application, consider the portfolios 1 and 2. Formally, investors will always prefer portfolio 1 to 2 if the mean-variance criteria are fulfilled:

\[ E[R_1] > E[R_2] \]  \hspace{1cm} (MVC 1)

and

\[ \sigma^2(R_1) \leq \sigma^2(R_2), \]  \hspace{1cm} (MVC 2)

or equivalently

\[ \sigma(R_1) \leq \sigma(R_2). \]

Of course portfolio 1 would also be preferred if
and
\[ E[R_1] \geq E[R_2] \quad \text{(MVC 1')} \]
\[ \sigma(R_1) < \sigma(R_2). \quad \text{(MVC 2')} \]
If (MVC 1) and (MVC 2) or (MVC 1') and (MVC 2') are true, portfolio 1 is said to dominate portfolio 2 (Cuthbertson & Nitzsche, 2004). As we will see later, MVA is the core in the derivation of the CAPM. The following mathematical arguments from utility theory will help to explain the motivation behind using it.

Consider a von Neumann-Morgenstern type utility function of an investor’s end of period wealth:
\[ u(W). \]
We evaluate this function at the expected end of period wealth by using a Taylor-series of the form
\[ \sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!} (x - a)^n, \]
where \( f \) is approximated for \( x \) near a point \( a \). As indicated we would like to evaluate \( f(x) = u(W) \) at \( a = E[W] \). Therefore the Taylor-expansion can be written as
\[ u(W) = u(E[W]) + u'(E[W])(W - E[W]) + \frac{1}{2} u''(E[W])(W - E[W])^2 + R_3, \quad (2.1) \]
where
\[ R_3 = \sum_{n=3}^{\infty} \frac{1}{n!} u^{(n)}(E[W])(W - E[W])^n. \]
If we instead consider expected values in (2.1), we can reduce the expression to the following more intuitively appealing version
\[ E[u(W)] = u(E[W]) + \frac{1}{2} u''(E[W])\sigma^2(W) + E[R_3], \quad (2.2) \]
where
\[ E[R_3] = \sum_{n=3}^{\infty} \frac{1}{n!} u^{(n)}(E[W]) m^n(W), \]
as the second term in (2.1) obviously cancels out. Note that \( m^n(W) \) indicates the \( n \)’th central moment of the end of period wealth (Huang & Litzenberger, 1988). If we assume that investors have strict concave utility functions, (2.2) supports MVA by first glance. It states a preference for wealth and an aversion to uncertainty in the form of variance of wealth. To see why, note that the second derivative (second term) must be negative if the utility curve is concave. Then, in order to maximize expected utility, investors will seek to maximize expected wealth while minimizing variance exactly as MVA suggests.

However, (2.2) also says that expected value and variance of wealth are not the only properties influencing expected utility. The term \( E[R_3] \) involves higher order moments indicating that
investors also care about skewness and kurtosis risk. It sounds reasonable that investors are interested in knowing the likelihood of extreme events for example.

Fortunately, under certain assumptions, MVA is still applicable. A straightforward implication of (2.2) is that if the utility function is quadratic, \( E[R^3] \) must equal zero. Then MVA is still useful, as investors will only care about minimizing the variance of their wealth while maximizing its the expected value. The reason is that \( E[R^3] \) only contains derivatives of a higher order than two. If the second or higher order derivative of a quadratic term is evaluated, the result is zero. Consider the general example

\[
u(W) = -aW^2 + bW + c,
\]

where \( a > 0 \) and \( a, b \) and \( c \) are exogenously given. The derivatives are \( u'(W) = -2aW + b \), \( u''(W) = -2a \), and \( u'''(W) = 0 \) for higher order derivatives. Since the second order derivative \( u''(W) = -2a \) is a constant, it makes sense substituting it into the expected value version of the Taylor expansion (2.2). We have

\[
E[u(W)] = u(E[W]) - aE[W]^2.
\]

Hence, if the utility function is quadratic, the first and second order moments (mean and variance) are the only distributional properties of interest to the investor. This is consistent with MVA. Mark that the term in the utility function (2.3) \(-aW^2\) must be negative thanks to \( a > 0 \) as defined above. This ensures that the utility curve is concave indicating that the investor is risk averse. The implication of the concave nature has, of course, carried through to (2.4), as \( E[u(W)] \) is a decreasing function of variance of wealth. It should be mentioned that this holds for arbitrary distributions. In other words, it does not matter which distribution wealth is assumed to follow. However, there is a highly doubtful assumption inherent in working with quadratic utility functions. The problem is that this type of function involves a point of satiation where the investor begins having preferences against more wealth. This point can be found as

\[
\frac{\partial u(W)}{\partial W} = -2aW + b = 0
\]

\(\leftrightarrow W = \frac{b}{2a}\)

In other words, there are diminishing returns to wealth in terms of utility. This seems unreasonable, as rational investors would always prefer more wealth to less. An alternative assumption that can support MVA is that wealth is, or equivalently returns are, normally distributed. This goes for arbitrary preferences (utility functions). In the normal distribution the moments of a higher order than two (i.e. those included in \( E[R^3] \)) can be expressed as functions of the first and second moment. Thus, in effect, investors only care about mean and variance under the normality assumption. One theoretical problem with this assumption is that the normal distribution has no lower bound. This is quite unrealistic in this setting. Sometimes you cannot lose more than you own, and there will surely always be limits to the amount of negative wealth (debt) an investor can have. Alternatively, one could point out the vast empirical evidence against the normal distribution as an approximation for financial data (Huang & Litzenberger, 1988; Campbell et al., 1997). We will treat distributional assumptions more thoroughly in chapter 3.

All in all, there are some theoretical arguments against MVA making sense. However, from now on, we shall assume that investors’ utility functions are concave and strictly increasing with no
point of satiation. Moreover, we assume that wealth is (or returns are) multivariate normally distributed along the lines of Lintner (1965 b). Thus, MVA can be applied in our analysis, as investors only care about minimizing variance and maximizing expected return in order to reach the highest possible utility-level. Section 2.1 can be summarized into the following simple results:

**Result 2.1 a:** For every given level of expected return, investors prefer the portfolio(s) with the lowest variance.

**Result 2.1 b:** For every given level of variance, investors prefer the portfolio(s) with the highest return.

### 2.2 The Minimum Variance Frontier

Let us extend the analysis from 2.1 to include every feasible portfolio. The ones that satisfy (MVC 1) and (MVC 2) when compared to every other portfolio are said to be *efficient*[^4]. Here efficient means that you cannot find a better portfolio in a mean-variance space. The set of feasible portfolios that meet the MVC constitute what is called the *efficient frontier*. First, we will illustrate the frontier graphically. Second, we will go through the computations necessary to identify the efficient portfolios. The analysis in this section is carried through assuming there is no risk-free rate. This assumption is relaxed in section 2.3.

To fully understand what lies behind the efficient frontier, let us first consider the graphical implications of combining two risky assets. When combining asset 1 and 2 in a portfolio, the expected return of that portfolio would be a weighted average of the two expected asset returns. We write

\[ E_P = X_1 E[R_1] + X_2 E[R_2], \]

where \( X_i \) indicates the proportion of wealth invested in asset \( i \) and \( X_1 + X_2 = 1 \). In the two-asset case, the variance can be found as

\[ \sigma_P^2 = X_1^2 \sigma_1^2 + X_2^2 \sigma_2^2 + 2X_1X_2 \rho_{1,2} \sigma_1 \sigma_2, \]

where \(-1 \leq \rho_{1,2} \leq 1\) as the correlation coefficient \( \rho_{1,2} \) is just a standardized version of the covariance \( \sigma_{12} \). Figure 2.1 demonstrates the risk-return frontier when combining portfolio 1 and 2 at different values of the correlation coefficient. The reason behind depicting the risk–return relationships in a return-standard deviation space will be obvious later.

The convex nature of the relationships indicates the incentive to hold a diversified portfolio. If the returns between asset 1 and 2 are perfectly correlated, making \( \rho = 1 \), the combinations will lie on a straight line connecting the two points. Thus, in terms of MVA, there are no gains from diversification when returns are perfectly correlated. However, as the correlation coefficient goes toward -1, the gain from diversification increases. This can be seen from the dashed line depicting a given level of expected return. Where the dashed line intersects with the different risk-return frontiers, there exists a certain combination of asset 1 and 2. It can be seen that for the same expected return, investors take on lower risk when the correlation decreases. This relationship is also evident in (2.6) where variance is increasing in \( \rho_{1,2} \). Thus, the risk-return frontiers with \( \rho < 1 \) must lie to the left of the straight line.

[^4]: Note that this assumption is purely for simplicity. Normality is not necessary, only sufficient. As we will see in chapter 3, other distributional assumptions are also consistent with MVA.

[^4]: Of course, two distinct efficient portfolios can have the same properties.
Figure 2.1 Risk-return frontiers

Source: (Cuthbertson & Nitzsche, 2004)

Figure 2.1 demonstrates that when asset returns do not experience exactly the same shocks, investors can gain from ‘not putting all their eggs in one basket’. It is easiest to grasp when thinking of the case where two assets have the same expected return and variance. If they are perfectly correlated, you do not gain from making some combination of them. However, if their correlation coefficient is sufficiently low, you might actually get a lower risk while the expected return is naturally given for any combination of the two (Cuthbertson & Nitzsche, 2004). In essence, we have demonstrated here is the importance of considering the correlation between securities when evaluating investment opportunities. It will be obvious later that the intuition regarding diversification and correlations we have just presented is closely related to the economic interpretation of the CAPM.

In order to get from figure 2.1 to variance analysis. Let there be a combination of two assets. A simple combination of many assets makes some portfolios available while combinations of these combinations make even more portfolios available. These operations are demonstrated in figure 2.2 a. The boundary of the feasible region some call the minimum variance frontier (Sharpe, 1964). Another more obvious name is the opportunity curve (Sharpe, 1964) as it contains the portfolios with the lowest variance for every level of expected return.

Figure 2.2 a: Risk-return frontiers asset case.

Source: Sharpe, 1970
In figure 2.2 a the curves connecting points 1, 2 and 3 represent the available combinations when working with two risky assets at a time. Consider the combination of 2 and 3 marked by Y. One could combine mix with portfolio 1, and a certain weighting would give the point Z. Continuing this exercise, it is easy to see why the feasible region must be entirely filled. This is illustrated in figure 2.2 b where the shaded area marks the available portfolios when evaluating all risky assets. It is the feasible region resulting from the infinite combinations available, and the blue line (both dashed and solid) is the minimum variance frontier (Sharpe, 1970).

Only a specific part of this region is interesting, however. The portfolios satisfying the mean-variance criteria can be interpreted as the portfolios farthest to the north-west. Recalling the MVC, it should be clear that the solid line on the upper part of the minimum variance curve consists of the efficient portfolios. This line is the efficient frontier. All other portfolios are inefficient including the dashed part of the frontier beginning at the global minimum variance portfolio. Every other portfolio in the feasible set give less return for the same levels of variance. The slope of the efficient frontier tells us something about how the investor can trade off expected return for risk. In other words, how much return you have to give up in order to reduce the risk of your portfolio with one unit. As one might expect, you have to give op more return to reduce risk when the risk is already low than if your portfolio is a very risky one. Consequently, there are diminishing marginal returns to risk in figure 2.2 b. We can summarize into the following result.

**Result 2.2 a:** For every level of portfolio risk as measured by its standard deviation, the portfolios with the highest expected returns are efficient.

By first mapping the entire minimum variance curve and then identifying the upper part of it, the efficient frontier can be traced out. In order to construct this curve, the portfolio weights of the single assets are needed. To find these, we must minimize the variance for every given level of expected return. This involves finding the optimal distribution of wealth between the individual stocks. This gives rise to the following minimization problem:

$$\min \sigma_p^2 = \sum_{i=1}^{N} (X_i^2 \sigma_i^2) + \sum_{i=1}^{N} \sum_{j=1}^{N} (X_i X_j \sigma_{ij})$$

(2.7)

When a set of optimal portfolio weights are obtained for some expected return, we have one point on the curve. Thus, one will have to redo this exercise until the whole frontier is traced out. Yet, to map out the efficient frontier it can be shown that we only need two distinct portfolios. This will now be proven.

In matrix notation the above expression for the portfolio variance is equivalent to

$$\sigma_p^2 = \omega' \Omega \omega.$$  

(2.8)

Let $\Omega$ be the $N \times N$ variance-covariance matrix of the $N$ assets in the portfolio. Let $\omega$ be the $N \times 1$-vector of portfolio weights on risky assets in $P$, and $\omega'$ be its transpose making it a $1 \times N$-vector. The link between the two expressions of the portfolio variance is presented in Appendix B. The minimization problem is

$$\min \frac{1}{2} \omega' \Omega \omega,$$

subject to
\[ \omega' \mu = \mu_p \]

and

\[ \omega' i = 1. \]

The \( \frac{1}{2} \) in front of the variance expression (2.9) is included to make the computations more neat. Let \( \mu \) be the \( N \times 1 \)-vector of expected returns of the risky assets and \( i \) be a \( N \times 1 \)-vector of ones. Furthermore, \( \mu_p \) equals the desired expected return of the optimal portfolio \( P \). Applying the

Lagrange multiplier method to the above problem, we start by forming the Lagrangian. We write

\[
\min_{\omega, \delta_1, \delta_2} L = \frac{1}{2} \omega' \Omega \omega + \delta_1 (\mu_p - \omega' \mu) + \delta_2 (1 - \omega' i),
\]

(2.10)

where \( \delta_1 \) and \( \delta_2 \) are the Lagrange multipliers. Then the first order conditions can be formed as

\[
\frac{\partial L}{\partial \omega} = \Omega \omega + \delta_1 (-\mu) + \delta_2 (-i) = 0,
\]

(2.11)

\[
\frac{\partial L}{\partial \delta_1} = \mu_p - \omega' \mu = 0,
\]

(2.12)

and

\[
\frac{\partial L}{\partial \delta_2} = 1 - \omega' i = 0.
\]

(2.13)

In the first condition the term \( \Omega \omega \) comes from the rule \( \partial (x'Ax)/\partial x = (A + A')x \). Note that the variance-covariance matrix \( \Omega \) is symmetric as \( \sigma_{ij} = \sigma_{ji} \) so the transpose of the matrix equals the matrix. Thus, \( (\Omega + \Omega') \omega = 1/2 \times 2 \Omega \omega = \Omega \omega \). In the other two terms containing \( \omega' \) we use the rule \( \partial (x'a)/\partial x = a \). The second and third conditions are obvious as \( \delta_1 \) and \( \delta_2 \) are just scalars. It can be seen in Appendix B that \( \omega' \Omega \omega \) yields a quadratic function since the portfolio weights are included twice in the expression. Also, it can be seen in Appendix A.1 that the quadratic terms are positive as variances of risky portfolios are, of course, strictly positive. Thus, the Lagrangian must be convex. This point is also evident in the first term in the formula for the portfolio variance using summation notation (2.7). In our setting, \( \Omega \) must be positive definite as \( \omega' \Omega \omega \) is the variance of a risky portfolio. Again, it follows that the Lagrangian is convex. When \( L \) is convex, it follows that the first order conditions above are both necessary and sufficient to find a global solution to the minimization problem. In order to find the solution, we have to solve (2.11)-(2.13) simultaneously. We start by solving for the portfolio weights \( \omega_p \) in (2.11). We write

\[
\Omega_p = -\delta_1 \Omega^{-1} (-\mu) - \delta_2 \Omega^{-1} (-i) \\
= \delta_1 (\Omega^{-1} \mu) + \delta_2 (\Omega^{-1} i).
\]

(2.14)

Multiplying by \( \mu' \) on both sides we get

\[
\mu' \omega_p = \delta_1 (\mu' \Omega^{-1} \mu) + \delta_2 (\mu' \Omega^{-1} i).
\]

(2.15)

Realizing that \( \mu' \omega_p = \omega' \mu \) we rearrange (2.12) and substitute

\[
\mu_p = \delta_1 (\mu' \Omega^{-1} \mu) + \delta_2 (\mu' \Omega^{-1} i).
\]

(2.16 a)
Multiplying both sides in (2.14) by $i$ we can rearrange (2.13) and substitute to get

$$1 = \delta^1(i \Omega^{-1} \mu) + \delta^2(i \Omega^{-1} t).$$

(2.16 b)

Now we have two equations with two unknowns $\delta^1$ and $\delta^2$. It is shown in Appendix A.2 that

$$\delta^1 = \frac{C \mu_p - A}{D}$$

(2.17 a)

and

$$\delta^2 = \frac{B - A \mu_p}{D},$$

(2.17 b)

where $A = i \Omega^{-1} \mu, B = \mu \Omega^{-1} \mu, C = i \Omega^{-1} t$ and $D = BC - A^2$. Substituting the above expressions for $\delta^1$ and $\delta^2$ in (2.14) it is derived in Appendix A.3 that

$$\omega_p = g + h\mu_p,$$

(2.18)

where

$$g = \frac{1}{D} \left[ B(\Omega^{-1} t) - A(\Omega^{-1} \mu) \right]$$

and

$$h = \frac{1}{D} \left[ C(\Omega^{-1} \mu) - A(\Omega^{-1} \mu) \right].$$

As we have used the sufficient conditions, the weights given by the vector $\omega_p$ result in a portfolio on the minimum variance frontier with expected return equal to $\mu_p$. A number of interesting conclusions can be drawn from this solution.

As we have solved the minimization problem for an arbitrary level of expected return, we can use (2.18) to find every frontier portfolio. In fact, every frontier portfolio can be constructed using two random frontier portfolios with different levels of expected return. To see why let $q$ be any portfolio on the minimum variance frontier. Since the two random frontier portfolios have different levels of expected return, there must exist a combination such that

$$\mu_q = X \mu_1 + (1 - X) \mu_2,$$

(2.19)

where $X$ is the weight placed on frontier portfolio 1. Assigning these weights to the weight-vectors of portfolio 1 and 2, and using (2.18) we have

$$X \omega_1 + (1 - X) \omega_2 = X[g + h \mu_1] + (1 - X)[g + h \mu_2]$$

$$= g + h[X \mu_1 + (1 - X) \mu_2]$$

$$= g + h \mu_q$$

(2.20)

The third equation follows from (2.19), and the last equation follows from (2.18). Remembering that (2.18) was the solution to the minimization problem, $\omega_q$ must be the weight-vector of a minimum variance frontier portfolio. As portfolios 1 and 2 could be any two distinct frontier portfolios, we get the following results:
Result 2.2 b:
Any minimum variance frontier portfolio is a linear combination of two other distinct frontier portfolios.

Result 2.2 b’:
The entire minimum variance frontier can be computed having two distinct frontier portfolios.

Instead of working with just two frontier portfolios, let us include a number of solutions to the minimization problem. Using (2.18) a linear combination of the weight-vectors of $k$ frontier portfolios can be written as

$$\sum_{i=1}^{k} X_i \omega_i = \sum_{i=1}^{k} X_i (g + h \mu_i),$$  \hspace{1cm} (2.21)

where

$$\sum_{i=1}^{k} X_i = 1.$$

As $g + h$ is the same for every $i$’th frontier portfolio we can write

$$\sum_{i=1}^{k} X_i \omega_i = g \sum_{i=1}^{k} X_i + h \sum_{i=1}^{k} X_i \mu_i = g + h \sum_{i=1}^{k} X_i \mu_i,$$  \hspace{1cm} (2.22)

When the weights on the different frontier portfolios sum to unity we have

$$\bar{\omega} = g + h \bar{\mu},$$  \hspace{1cm} (2.23)

where

$$\bar{\omega} = \sum_{i=1}^{k} X_i \omega_i$$

and

$$\bar{\mu} = \sum_{i=1}^{k} X_i \mu_i.$$

Comparing (2.23) with (2.18), it is evident that any linear combination $\bar{\omega}$ of solutions to the minimization problem is itself a solution to the problem. Thus we get the following more powerful result:

Result 2.2 c:
Any linear combination of minimum variance frontier portfolios is itself a minimum variance frontier portfolio.

Recall from figure 2.2 b that the part of the minimum variance frontier below the global minimum variance portfolio is inefficient. In contrast, the portfolios on the upper part of the frontier are efficient. Thus, every portfolio with expected return above that of the global minimum variance frontier must be efficient. Let and $X_{e1}, X_{e2}, \ldots, X_{ek}$ be the nonnegative weights assigned to $k$ efficient portfolios. If the considered portfolios are all efficient then $\mu_i \geq \mu_{\text{min}}$ for $i = 1, 2, \ldots, k$.
where \( \mu_{min} \) is the expected return for the global minimum variance portfolio. When we apply these constraints to the solution in (2.23), we have

\[
\bar{\mu} = \sum_{i=1}^{k} X_i \mu_{i} \ge \sum_{i=1}^{k} X_i \mu_{min}.
\]  

(2.24)

As the \( X_i \) must sum to unity we get

\[
\bar{\mu} \ge \mu_{min}.
\]  

(2.25)

So when only efficient portfolios are combined to form the weights \( \bar{\omega} \), the expected return \( \bar{\mu} \) of this mix is above that of the global minimum variance portfolio. This can be summarized in the following result:

**Result 2.2 d:** Any convex\(^5\) combination of efficient portfolios is itself an efficient portfolio.

We now claim that the market portfolio is a frontier portfolio. To prove this, let us first consider the definition of the market portfolio. The aggregated wealth of the economy is

\[
W_M \equiv \sum_{i=1}^{l} W_i,
\]  

(2.26)

where \( l \) denotes the total number of individuals in the economy. In equilibrium it follows that

\[
\sum_{i=1}^{l} W_i X_{ij} = W_M X_{mj},
\]  

(2.27)

where \( X_{ij} \) is individual \( i \)’s portfolio weight on security \( j \). Then \( X_{mj} \) must be the \( j \)’th security’s share of the total value of all assets. Rearranging we have

\[
\sum_{i=1}^{l} \frac{W_i}{W_M} X_{ij} = X_{mj}.
\]  

(2.28)

If we assume that every individual has positive wealth, (2.28) says that the market portfolio weights are a convex combination of every individual’s portfolio weights.

**Result 2.2 e:** The market portfolio is a convex combination of the individual portfolios.

Recall that investors only choose efficient portfolios when applying MVA. Combining 2.2 d and 2.2 e we obtain a very important result (Huang & Litzenberger, 1988):

**Result 2.2 f:** The market portfolio is mean-variance efficient.

This is result should be noted carefully, as it will be our key testable implication of the CAPM. We return to this point in chapter 3.

---

\(^5\) A convex combination is linear combination where the coefficients are nonnegative and sum to unity.


### 2.3 The Capital Market Line

Figure 2.2 b depicts the efficient frontier when only risky assets are available. Now, we include an asset with a riskless rate also known as a pure interest rate. Let us consider a combination of any risky portfolio and the risk-free asset. Modifying (2.5) a bit, the expected return on this combination would be:

\[ E[R_p] = (1 - X)R_f + X E[R_T] \]  

(2.29 a)

or

\[ E[R_p] = R_f + X(E[R_T] - R_f), \]  

(2.29 b)

subject to \( 0 \leq X < \infty \). Note that \( X \) is the proportion of wealth invested in the risky portfolio, and \( (1-X) \) is the proportion invested at the risk-free rate \( R_f \). Here \( X = 1 \) implies that the investor holds all his wealth in the risky portfolio. Conversely, \( X < 1 \) means that the investor lends some of his wealth receiving the risk-free rate. Finally, \( X > 1 \) implies a leveraged position where he borrows money at the risk-free rate as \( (1-X) < 0 \). He then invests the proceeds from the loan plus his wealth in stocks and other risky assets.

The variance of a variable \( Y \) can be written as \( \sigma_y^2 = E[(Y - E[Y])^2] \). As the risk-free rate is a constant its expected value \( E[Y] \) must be the same as the realization \( Y \). Thus, its variance must be zero. Equivalently, the covariance between a constant and any variable must be zero. This can easily be seen from the formula for the covariance \( \sigma_{YZ} = E[(Y - E[Y])(Z - E[Z])] \) where the same logic used above applies. Recalling (2.6), we can use these simple observations for determining the variance of the portfolio above to get

\[ \sigma_p^2 = X^2 \sigma_r^2 \]  

(2.30 a)

or

\[ \sigma_p = X \sigma_r. \]  

(2.30 b)

As we are interested in placing this portfolio in a mean–standard deviation space, we need an expression containing these to parameters. We therefore rearrange (2.9) and substitute it into (2.7 b) and get

\[ E[R_p] = R_f + \theta \sigma_p, \]  

(2.31)

where

\[ \theta = \frac{(E[R_T] - R_f)}{\sigma_T}. \]

The expression in (2.31) says that the net expected rate of return on the total investment is linearly related to the risk of the net investment. This expression can be used for any portfolio of risky assets and is named the market opportunity line. In figure 2.3 this is depicted as the line \( R_T A \) for some random risky portfolio. As opposed to the minimum variance frontier in figure 2.2 b, the market opportunity line expresses a constant tradeoff between risk and expected return. Therefore, the price of risk reduction is independent of the risk level of a given position.

Obviously \( \theta \), also known as Sharpe ratio, measures the slope of the line. This ratio is sometimes referred to as the price of risk for efficient portfolios. The intercept \( R_f \) is equivalently referred to as the price of time. The price of risk in particular is somehow misleading since one would never pay
for risk. A more precise name, according to Sharpe (1970), is therefore the price of risk reduction. It is a measure of how much net expected return you have to give up to reduce the risk by one unit. Likewise, he argues that the intercept could be named the price of immediate consumption.

**Figure 2.3 The optimal portfolio choice in a mean-standard deviation space.**

![Portfolio Choice Diagram](image)

Source: Lintner, 1965

According to (2.29 b), no matter which risky portfolio an investor chooses to combine with the risk-free asset, he can reach any desired net expected return of his total investment. He just has to leverage his investment in the risky asset enough by borrowing money. Yet, (2.30 b) reveals a serious drawback to leveraging a random portfolio: the risk of the total investment increases proportionately with the magnitude of the leverage $X$.

The investor has to deal with two questions:

1. Which risky portfolio (i.e. mix of risky assets) should he choose?
2. How intensively will he use it? (That is, how much of his wealth will he place in risky investments as measured by $X$?)

*Regardless of his utility function (subjective preferences), a risk averse investor will always choose the risky portfolio with the highest $\theta$. To see why, recall from above that he can reach an expected total net return $E[R_p]$ as high as he may desire with all portfolios. By choosing the risky portfolio with the maximum $\theta$-value, he minimizes the standard deviation and variance of his position for any level of expected return. Intuitively, this way he gets more return per risk-unit so he only has to take on minimal risk in order to reach a certain level of expected return.*

**Result 2.3 a:** For each investor, the optimal mix of risky assets is independent of the ratio of his total wealth invested in it.
This is equivalent to saying that question 1 can be solved independently of question 2\(^6\). This is known as the separation theorem (Lintner, 1965). The name stems from the phenomenon that the computation of the optimal mix of risk securities is separated from investors’ attitudes towards risk.

In figure 2.3, maximizing $\theta$ means choosing the market opportunity line with the steepest slope. Note that, as indicated by (2.31), every such line will always be straight. The feasible set can be seen as a constraint to the maximization problem. Obviously, the line with the maximal slope is tangent to the efficient frontier. The point is marked by a solid blue spot in figure 2.3. Under the assumptions listed in the beginning of the chapter, this tangent portfolio contains the optimal mix of risky assets for every investor. The line it forms is commonly known as the Capital Market Line (CML). The portfolios on this line are different mixes of the tangency portfolio of risky assets and the risk-free asset. As they lie on the same line, these portfolio must all have the maximal $\theta$. It is impossible to find a portfolio that would dominate any of them in terms of the mean-variance criteria. Thus, when including a risk-free asset, the CML constitutes the new efficient frontier. For every given level of risk, the portfolios on it have the maximum expected return making them mean-variance efficient. If it was not possible to leverage the tangency portfolio, the new efficient frontier would instead include the CML up to the point of tangency and thereafter the curved minimum variance frontier.

As illustrated in figure 2.3, the only difference between investors is the proportion of their wealth placed in the tangency portfolio of risky assets. Consider investor 1 with indifference curves $U_{11}$, $U_{12}$, and $U_{13}$. Investor 1 is relatively risk averse, so he maximizes his utility where $U_{11}$ is tangent to the CML in point $i_1$. As a consequence, investor 1 chooses to lend some of his money resulting in less risk and expected return than that of a pure risk asset portfolio. As another example, we have investor 2 with indifference curves $U_{21}$, $U_{22}$, and $U_{23}$. He is a lot less risk averse than investor 1. Therefore, he borrows money in order to invest more than his wealth in the fixed mix of risk assets. As long as investors like wealth and dislike risk, some combination of the tangency portfolio and the risk-free asset will always maximize their utility.

The implication of the separation theorem is important when we want to describe the tangency portfolio. Recall the assumptions set up in the beginning of the chapter. Two of them are of particular interest when we want to make conclusions about the tangency portfolio. The first is homogeneity of expectations, and the second is unlimited lending and borrowing at the risk-free rate. When everyone have the same distributional expectations, and everyone can use the same risk-free rate, every investor must face the same minimum variance frontier and opportunity market line as illustrated in figure 2.3. Thus, every investor faces the same tangency portfolio. Then, as the separation theorem suggests, the optimal portfolio is independent of individual preferences. When all investors assign the same weight in their risk portfolios to any given security, they must all hold the market portfolio (Sharpe, 1970).

**Result 2.3 b:** When the market is in equilibrium, the tangency portfolio is the market portfolio.

In section 2.1 we made the assumption that investors’ utility curves were strictly increasing. It can be seen in figure 2.3 that they consequently always choose efficient portfolios\(^7\). As the market

---

\(^6\) Yet, question 2 is dependent on question 1 as question 2 is affected by the shape of the investor’s utility function (how risk averse he is).
portfolio is a convex combination of every individual portfolio, the market portfolio must also be mean-variance efficient when involving a risk-free asset. Alternatively, one could argue that if the market portfolio is the tangency portfolio, it must lie on the CML and thus be efficient. Then, the slope in (2.31) can be written as:

$$\theta = \frac{(E[R_m] - R_f)}{\sigma_m}.$$  \hspace{1cm} (2.32)

Note, that if we relaxed the assumption regarding unlimited borrowing, the tangency portfolio would not necessarily be the optimal risk-choice for everyone. Recall from above, that without the possibility to leverage, the part of the efficient frontier to the right of the tangency point would comprise some part of the curved minimum variance frontier. Now imagine an investor with preferences like investor 2 from the example. The point where one of his utility curves would be tangent to this efficient frontier could be marked by the non-solid blue spot in figure 2.3. This optimal risky portfolio is different from that of investor 1 as he holds some proportion of the tangency portfolio. Thus, every investor would not necessarily assign the same portfolio weights to a given risky asset (Sharpe, 1970). As demonstrated, the conclusions presented above rely heavily on the assumptions made in the start of the chapter.

The CML only describes efficient portfolios. It does not relate any expected return to risk. However, the CAPM does. Let us therefore use what we have learned about mean-variance analysis and the CML to derive the CAPM. We will do it in three different ways each representing an alternative approach. Of course, in some sense they are all the same, but the different approaches will each contribute to a more thorough understanding of where the CAPM comes from.

### 2.4 Derivations of the CAPM

As indicated in the introduction, William F. Sharpe and John V. Lintner are commonly said to be the architects of the CAPM model while Markowitz laid the groundwork. Sharpe, in 1964, and Lintner, in 1965, independently derived the model, and their approaches will both be presented here. We start with Sharpe’s as it is rather straightforward.

#### 2.4.1 Sharpe’s derivation

Consider a combination of the market portfolio and any portfolio $i$. Denote this portfolio $Z$. Then according to (2.5) and (2.6) the expected return on this combination is

$$E[R_Z] = (1 - X)E[R_m] + XE[R_i],$$  \hspace{1cm} 2.33

and the standard deviation is

$$\sigma_Z = [(1 - X)^2 \sigma_m^2 + X^2 \sigma_i^2 + 2X(1 - X)\rho_{mi}\sigma_m\sigma_i]^{1/2}. $$  \hspace{1cm} 2.34

At $X = 1$ only asset $i$ is held while $X = 0$ implies that only the market portfolio is held. Points $i$ and $M$ in figure 2.4 illustrate the two cases. Note however, that at $X = 0$ some wealth is still placed in asset $i$ as it is a part of the market. Thus, at $X = 0$ the weight placed on asset $i$ corresponds to its values proportion of the total market wealth. $M’$ indicates the case where $i$ is not held at all making $X < 0$ (Sharpe, 1964). It should be marked that this combination curve does not intersect the minimum variance frontier. By definition, when there is no riskless asset all efficient portfolios lie

---

7 In the mean-standard deviation space depicted in figure 2.3 the indifference curves are convex to the “risk axis” which is consistent with concave utility functions (Cuthbertson and Nitzsche, 2004).
on the curved minimum variance frontier, so the combinations between \( m \) and \( i \) cannot dominate them.

**Figure 2.4 Combining the market portfolio with some random risk asset.**

![Diagram of market portfolio and random risk asset combinations](image)

Source: (Cuthbertson & Nitzsche, 2004)

As indicated above, we want an expression for the expected value of any asset. In order to obtain this we investigate the curve comprising the combinations of \( i \) and the market portfolio at point \( M \). Note that \( i \) does not have to be efficient. To find the slope of the curve we differentiate \([R_Z] \) with respect to \( \sigma_Z \). As \( \sigma_Z \) is not in the expression for \([R_Z] \), we use the fact that

\[
\frac{\partial[R_Z]}{\partial \sigma_Z} = \frac{\partial[R_Z]}{\partial \sigma_Z}. \tag{2.35}
\]

First we take the derivative of the total expected return on portfolio \( Z \) with respect to the weight on asset \( i \):

\[
\frac{\partial[R_Z]}{\partial X} = -E[R_m] + E[R_i]. \tag{2.36}
\]

Then we take the derivative of the standard deviation of portfolio \( Z \) with respect to the weight on asset \( i \) by using the chain rule:

\[
\frac{\partial \sigma_Z}{\partial X} = \frac{1}{2} \left[ (1 - X)^2 \sigma_m^2 + X^2 \sigma_i^2 + 2X(1 - X)\rho_{mi} \sigma_m \sigma_i \right]^{1/2} \\
\times \left[ -2(1 - X)\sigma_m^2 + 2X\sigma_i^2 + 2(1 - 2X)\rho_{mi} \sigma_m \sigma_i \right] \\
= \frac{1}{\sigma_Z} \left[ -(1 - X)\sigma_m^2 + X\sigma_i^2 + (1 - 2X)\rho_{mi} \sigma_m \sigma_i \right]. \tag{2.37}
\]

\( \textsuperscript{8} \) We only need the derivative with respect to the weight on asset \( i \) as the weight on the market portfolio has been expressed in terms of the weight on asset \( i \).
At point M we recall that \( X = 0 \). At this point \( \sigma_Z = \sigma_m \) since we only hold the market portfolio. Furthermore, we use the fact that \( \sigma_{mi} = \rho_{mi} \sigma_m \sigma_i \). Now (2.37) reduces to

\[
\frac{\partial \sigma_Z}{\partial X} = \frac{-\sigma_m^2 + \sigma_{mi}}{\sigma_m}.
\] (2.38)

Rearranging (2.38) and (2.36) and inserting into (2.35) we get

\[
\frac{\partial [R_Z]}{\partial \sigma_Z} = \frac{(E[R_i] - E[R_m]) \sigma_m}{\sigma_m - \frac{\sigma_m^2}{\sigma_M}} = S_Z,
\] (2.39)

where \( S_Z \) is the slope of the combination curve at point M. In equilibrium, this curve must be tangent to the CML which makes their slopes equal. In other words, in equilibrium, the tradeoff between expected return and risk in the combination of asset \( i \) and the market must equal the tradeoff in the aggregated capital market (Sharpe, 1970). By equating the two slopes, we can rearrange to get the traditional equilibrium model known as the CAPM. We have

\[
S_Z = 0.
\] (2.40)

Inserting the expressions we get

\[
\frac{(E[R_i] - E[R_m]) \sigma_M}{\sigma_m} = \frac{(E[R_m] - R_f)}{\sigma_m - \frac{\sigma_m^2}{\sigma_M}} + E[R_m]
\]

\[
\Leftrightarrow E[R_i] = \frac{\sigma_{mi} - \frac{\sigma_m^2}{\sigma_M}}{(E[R_m] - R_f)(\sigma_{mi} - \frac{\sigma_m^2}{\sigma_M}) + E[R_m]}
\]

\[
= \frac{(E[R_m] - R_f) \sigma_{mi} - \frac{\sigma_m^2 (E[R_m] - R_f)}{\sigma_m} + E[R_m]}{\sigma_m}
\]

\[
= R_f + \beta_i (E[R_m] - R_f),
\] (2.41)

where

\[
\beta_i = \frac{\sigma_{mi}}{\sigma_m^2}.
\]

This is the traditional CAPM model derived by Sharpe in 1964. It says that the expected return of asset \( i \) equals the risk-free rate plus a reward for bearing risk. By examining the definition of beta, we get the next result.

**Result 2.4 a:** The covariance of returns between the market and asset \( i \) is the only individual property that influences the excess return of asset \( i \) above “the sure thing”.

An interesting conclusion can be drawn from the fact that we derived that CAPM where the pure market portfolio was held. An interpretation of \( X = 0 \) is that asset \( i \) is neutrally weighted. In other words, its weight in portfolio \( Z \) equals its proportion of the market. Thus, CAPM expresses the expected return asset \( i \) must yield in order to deserve a neutral weight in one's portfolio.

**2.4.2 Lintner’s derivation**

In 1965 (b) Lintner took a different, but similar, approach to deriving the CAPM. He based his analysis on the separation theorem and consequently maximizing the slope of the CML. He then
used the first order conditions from this maximization problem to derive an equilibrium model. The following exposition is adapted from his 1965-paper.

In figure 2.3 investors had to find the risky portfolio with the maximum \( \theta \)-value in order to reach the highest possible utility curve. As we saw, this portfolio would form the optimal market opportunity line regardless of his attitude towards risk. The maximization problem is the following

\[
\max \theta = \frac{(E[R_T] - R_f)}{\sigma_T},
\]

subject to

\[
\sum_{i=1}^{N} X_i = 1,
\]

where \( X_i \) is the weight assigned to asset \( i \) in the optimal portfolio \( T \). \( E[R_T] \) is just a weighted average of the expected returns of the assets in the portfolio. And since the weights must sum to unity we can write

\[
R_f = \left( \sum_{i=1}^{N} X_i \right) R_f.
\]

Using (2.7) as the formula for the variance of portfolio \( T \), we can rewrite the equation for the slope as

\[
\theta = \frac{\sum_{i=1}^{N} X_i (E[R_i] - R_f)}{\left[ \sum_{i=1}^{N} (X_i^2 \sigma_i^2) + \sum_{i=1}^{N} \sum_{j=1, j \neq i}^{N} (X_i X_j \sigma_{ij}) \right]^{1/2}}.
\]

In this maximization problem there are \( N \) weights and therefore \( N \) variables we need to solve for in order to construct the optimal portfolio of risky assets. To solve the problem we take the partial derivative with respect to each weight and set it equal to zero:

\[
\frac{\partial \theta}{\partial X_1} = 0, \frac{\partial \theta}{\partial X_2} = 0, \ldots, \frac{\partial \theta}{\partial X_N} = 0.
\]

This system of equations must be solved simultaneously. The derivative with respect to the \( k \)’th weight can be determined using the quotient rule. We have

\[
\frac{\partial \theta}{\partial X_k} = \frac{\frac{\partial G(X)}{\partial X_k} H(X) - G(X) \frac{\partial H(X)}{\partial X_k}}{[H(X)]^2},
\]

where

\[
G(X) = (E[R_T] - R_f) = \sum_{i=1}^{N} X_i (E[R_i] - R_f)
\]
Taking the derivative of (2.46) with respect to $X_k$ we get

$$\frac{\partial G(X)}{\partial X_k} = \mathbb{E}[R_k] - R_f,$$

as all the other terms are constants when evaluating $X_k$. Taking the derivative of (2.47) with respect to $X_k$ we need the chain rule. The result is

$$\frac{\partial H(X)}{\partial X_k} = \frac{1}{2} \left[ \sigma_T^2 \right]^{-1/2} \left[ 2X_k \sigma_k^2 + 2 \sum_{j=1}^{N} X_j \sigma_{jk} \right],$$

where the expression $2X_k \sigma_k^2$ in the first equality also comes from the fact that all the other terms are constants when evaluating $X_k$. The computation behind $2 \sum_{j=1,j \neq k}^{N} X_j \sigma_{jk}$ is not so straightforward.

Note that the anti-derivative of this term is $\sum_{i=1}^{N} \sum_{j=1,j \neq i}^{N} (X_i X_j \sigma_{ij})$. There are only two cases with $X_k$ in that expression, namely when $i = k$ and when $j = k$. They are $\sum_{j=1,j \neq k}^{N} X_k X_j \sigma_{kj}$ and $\sum_{j=1,j \neq k}^{N} X_i X_k \sigma_{ik}$. As $i$ is $j$ mere notation the derivatives of these two cases must be equal, and thus the term above is obtained (Elton and Gruber, 1995). Inserting (2.49) and (2.48) into (2.45) and setting it equal to zero yields

$$\frac{\partial \theta}{\partial X_k} = \frac{(\mathbb{E}[R_k] - R_f) \sigma_T - (\mathbb{E}[R_T] - R_f) \sigma_T^{-1} \left[ X_k \sigma_k^2 + \sum_{j=1}^{N} X_j \sigma_{jk} \right]}{\sigma_T^2},$$

where

$$\gamma = \frac{(\mathbb{E}[R_T] - R_f) \sigma_T^{-2}}{\mathbb{E}[R_m] - R_f} = \frac{(\mathbb{E}[R_T] - R_f)}{\sigma_T^2}.$$
Similarly to the interpretation of $\theta$ the fraction $\gamma$ can be understood as the market price of risk. In order for $\partial \theta / \partial X_k$ to equal zero the numerator has to equal zero. Thus, in (2.50) we only have to solve

$$
(E[R_k] - R_f) = \gamma \left[ X_k \sigma_k^2 + \sum_{j=1}^{N} X_j \sigma_{jk} \right].
$$

(2.52)

The term in the brackets on the right side of the equation is actually the covariance between the returns on asset $k$ and the portfolio with weights $X_j$. We have

$$
X_k \sigma_k^2 + \sum_{j=1,j \neq k}^{N} X_j \sigma_{jk} = \sum_{j=1}^{N} X_j \sigma_{jk} = \sum_{j=1}^{N} X_j E[(R_j - E[R_j])(R_k - E[R_k])]
$$

$$
= E \left[ \left( \sum_{j=1}^{N} X_j R_j - \sum_{j=1}^{N} X_j E[R_j] \right) (R_k - E[R_k]) \right]
$$

$$
= E[(R_m - E[R_m])(R_k - E[R_k])] = \sigma_{mk},
$$

(2.53)

where the fourth equality follows from the fact that $X_j$ must equal the market proportions since the optimal portfolio $T$ is the market portfolio (Cuthbertson & Nitzsche, 2004). Substituting (2.53) and (2.51) into (2.52) yields

$$
(E[R_k] - R_f) = \frac{(E[R_m] - R_f)}{\sigma_m^2} \sigma_{mk}
$$

$$
\leftrightarrow E[R_k] = R_f + \beta_k (E[R_m] - R_f),
$$

(2.54)

where

$$
\beta_k = \frac{\sigma_{mk}}{\sigma_m^2}.
$$

This is exactly the same model we derived before using Sharpe’s approach. We have now shown that the CAPM lies implicit in the first order conditions for finding the optimal risky portfolio. In his derivations Lintner pointed out that the optimal weight placed on asset $k$ depends on its variance and covariances and its expected return. As the optimal weights are the market portfolio weights, this indicates that the CAPM illustrates the required return on asset $k$ for it to be weighted in accordance with its market value. The expected return must reflect the asset’s contribution to the total portfolio risk. This is a similar interpretation to the neutral weight condition discussed above (Lintner, 1965 b).

### 2.4.3 Derivation by solving a minimization problem

We just shown that the CAPM can be derived by maximizing $\theta$. By obtaining the maximum price of risk, one can achieve the highest possible expected return for a given variance. Now will use a perfectly equivalent method as we turn the problem around and minimize the variance for a given level of expected return.

---

9 See the earlier discussion about the relevance of this interpretation.
This minimization problem has already been outlined under the treatment of the minimum variance frontier only without a risk-free asset. We wrote the formula for the variance of portfolio \( \sigma_P^2 \) as

\[
\sigma_P^2 = \omega' \Omega \omega.
\]

Thus, the minimization problem is

\[
\min_{\omega} \omega' \Omega \omega,
\]

subject to

\[
\omega' \mu + (1 - \omega' \mathbf{i}) R_f = \mu_p.
\]

This constraint is equivalent to the expression for the expected return we used to derive the CML. Wealth is once again split between a portfolio of risky assets and the risk-free rate. As in section 2.2, let \( \mu \) be the \( N \times 1 \)-vector of expected returns of the risky assets and \( \mathbf{i} \) be a \( N \times 1 \)-vector of ones. Also, \( \omega \) is a \( N \times 1 \)-vector of weights on the risky assets. As previously, \( R_f \) is the risk-free rate, and \( \mu_p \) is the sum of two scalars equaling the desired expected return of the optimal portfolio \( P \). For this given level of expected return we wish to minimize the variance. To solve the programming problem we use the Lagrange multiplier method. We thus form the Lagrangian

\[
\min_{\omega, \delta} \mathcal{L} = \omega' \Omega \omega + \delta (\mu_p - \omega' \mu - (1 - \omega' \mathbf{i}) R_f),
\]

where the first order conditions are

\[
\frac{\partial \mathcal{L}}{\partial \omega} = 2 \Omega \omega + \delta (-\mu + \mathbf{i} R_f) = 0 \tag{2.59 a}
\]

and

\[
\frac{\partial \mathcal{L}}{\partial \delta} = \mu_p - \omega' \mu - (1 - \omega' \mathbf{i}) R_f = 0. \tag{2.60 a}
\]

The calculus rules applied here are of course the same as for the solution without a risk-free asset in section 2.2. Let us first rearrange (2.60 a) to get

\[
\mu_p - R_f = \omega' \mu - \omega' \mathbf{i} R_f. \tag{2.60 b}
\]

Furthermore, (2.59 a) can be written as

\[
2 \Omega \omega = -\delta (-\mu + \mathbf{i} R_f)
\]

\[
\Leftrightarrow 2 \Omega \omega \delta = \mu - \mathbf{i} R_f
\]

\[
\Leftrightarrow \frac{2 \omega' \Omega \omega}{\delta} = \omega' \mu - \omega' \mathbf{i} R_f, \tag{2.59 b}
\]

where the last expression comes from multiplying by \( \omega' \) on both sides. We have to solve the two conditions simultaneously. We thus substitute (2.60 b) into the right hand side of (2.59 b) yielding

\[
\frac{2 \omega' \Omega \omega}{\delta} = \mu_p - R_f.
\]

\[
\Leftrightarrow \delta = \frac{2 \omega' \Omega \omega}{\mu_p - R_f}. \tag{2.61}
\]
Inserting (2.61) in (2.59 a) the following relationship appears (Skovmand, 2011)

\[
\frac{2\Omega \omega}{2 \omega \Omega \omega} = \mu - tR_f \\
\mu_p - R_f \iff \mu - tR_f = \frac{\Omega \omega}{\omega \Omega \omega} (\mu_p - R_f).
\] (2.62)

We already know that \( \omega' \Omega \omega \) is the portfolio variance. The interpretation of \( \Omega \omega \) is actually also very straightforward. As \( \Omega \) is a \( N \times N \)-matrix and \( \omega \) is a \( N \times 1 \)-vector their product exists and is a \( N \times 1 \)-vector\(^{10}\)

\[
\Omega \omega = \begin{bmatrix}
\sigma_{11} & \sigma_{12} & \cdots & \sigma_{1N} \\
\sigma_{21} & \sigma_{22} & \cdots & \sigma_{2N} \\
\vdots & \vdots & \ddots & \vdots \\
\sigma_{N1} & \sigma_{N2} & \cdots & \sigma_{NN}
\end{bmatrix}
\begin{bmatrix}
X_1 \\
X_2 \\
\vdots \\
X_N
\end{bmatrix}
\]

\[
= \begin{bmatrix}
(\sigma_{11}X_1 + \sigma_{12}X_2 + \cdots + \sigma_{1N}X_N) \\
(\sigma_{21}X_1 + \sigma_{22}X_2 + \cdots + \sigma_{2N}X_N) \\
\vdots \\
(\sigma_{N1}X_1 + \sigma_{N2}X_2 + \cdots + \sigma_{NN}X_N)
\end{bmatrix}
\] (2.63)

Each of the entries in (2.63) equals the following expression from the Sharpe’s derivation in section 2.4.1:

\[
X_k \sigma_k^2 + \sum_{j=1}^{N} X_j \sigma_{jk}.
\]

In that derivation, we showed that this is in fact the covariance between the expected returns of asset \( k \) and the portfolio with weights \( X_j \). To see why, notice in the first entry of \( \Omega \omega \) that \( \sigma_{11}X_1 = X_1 \sigma_1^2 \). This is the first term in the above formula for \( k = 1 \) while the rest concerns the covariances. Thus, \( \Omega \omega \) is a \( N \times 1 \)-vector of the covariances between the expected returns on the \( N \) risky assets and portfolio \( P \). Consequently we can write

\[
\frac{\Omega \omega}{\omega' \Omega \omega} = \beta_P,
\] (2.64)

where \( \beta_P \) is a \( N \times 1 \)-vector in which the entries in \( \Omega \omega \) are multiplied by the inverse of the scalar \( \omega' \Omega \omega \). Noting that the \( k \)’th entry equals \( \sigma_{pk} \) and \( \omega' \Omega \omega \) equals \( \sigma_p^2 \) the vector becomes

\[
\beta_P = \begin{bmatrix}
\sigma_{p1}/\sigma_p^2 \\
(\sigma_{p2}/\sigma_p^2) \\
\vdots \\
(\sigma_{PN}/\sigma_p^2)
\end{bmatrix}.
\] (2.65)

Thus, \( \beta_P \) is a vector of the \( N \) asset’s betas. Inserting in (2.62) we obtain

\(^{10}\) See appendix B for more details regarding the product of two matrices.
\[ \mathbf{\mu} - \mu \mathbf{R}_f = \mathbf{\beta}_p (\mu_p - \mathbf{R}_f) . \]  

(2.66)

The optimal portfolio \( P \) is not necessarily the market portfolio. However, we do know that it is efficient as it, for a given expected return \( \mu_p \), has the minimum variance of all combinations between risky assets and the risk-free asset. Yet, the CAPM explicitly involves the market portfolio. Consequently, we must impose the restriction on the solution to the minimization problem \( \mathbf{\omega}_p \) that the weights must sum to unity. In other words, the risk-free asset is not held. When this restriction is imposed we get the tangency portfolio according to figure 2.3. As argued before, when everybody have the same expectations and face the same, constant risk-free rate, the tangency portfolio must equal the market portfolio. Thus, portfolio \( P \) is the market portfolio. Rearranging (2.66) we get

\[
\begin{bmatrix}
E[R_1] \\
E[R_2] \\
\vdots \\
E[R_N]
\end{bmatrix}
= \begin{bmatrix}
1R_f \\
1R_f \\
\vdots \\
1R_f
\end{bmatrix} + \begin{bmatrix}
\beta_1(\mu_m - R_f) \\
\beta_2(\mu_m - R_f) \\
\vdots \\
\beta_N(\mu_m - R_f)
\end{bmatrix} = \begin{bmatrix}
R_f + \beta_1(\mu_m - R_f) \\
R_f + \beta_2(\mu_m - R_f) \\
\vdots \\
R_f + \beta_N(\mu_m - R_f)
\end{bmatrix},
\]

(2.67)

where the first equality comes from multiplying the scalars in into the vectors, and the second equality comes from adding the two vectors entry by entry. The difference between this equation and the former CAPM expressions is that it, due to the vectors, involves the entire set of equations relating expected return to beta. Above, we indicated that the relation expressed in the CAPM in fact applies for every efficient portfolio. On the other hand, as \( \mathbf{\omega}_p \) ensures that we are on the CML by minimizing portfolio variance given \( \mu_p \), it also only applies for efficient portfolios. As we will see in chapter 3, this implies that one can test the CAPM by examining whether the market portfolio is efficient.

2.5 The Security Market Line

The CAPM shows a linear relationship between expected return and beta. This linear relationship can be graphically presented as the security market line (SML), and if the CAPM holds, all assets lie on the SML. Recalling the Sharpe-Lintner CAPM-model it should be clear that if \( \beta_k = 1 \) the expected return of asset \( k \) equals the expected return of the market portfolio. This is illustrated in figure 2.5.

The SML indicates why the CML was not sufficient to price assets. Most portfolios and securities lie scattered somewhere below the CML as they are undiversified and thus inefficient. There was no theoretical consistent relationship between variance/standard deviation and expected return. Consequently, we needed a new risk indicator. Under the derivations of the CAPM, we found that beta would be a suitable choice. If the CAPM holds, we get the following result (Sharpe, 1964):

**Result 2.5 a:** Beta can completely explain the difference in expected returns between any assets.
One implementation of the SML is in the identification of over- and underpriced securities. Note that if a stock lies over the SML, it gives an abnormal expected return compared to what its beta would predict. This must mean this its current price is too low compared to the end price in one period. When the market finds out, demand will drive up the stock’s quote until the expected return is consistent with its beta. An underpriced asset in terms of the CAPM is sometimes referred to as having a positive Jensen’s alpha. It should be noted that this simple stock picking strategy relies on the CAPM to hold and that the betas are correctly estimated (Cuthbertson & Nitzsche, 2004).

Another implementation of the CAPM is in the estimation of a company’s cost of capital (the WACC). This can be used in a variety of applications such as project evaluation, budgeting etc.

2.5.1 Beta as a risk measure
To evaluate beta as a risk measure let us first review the definition:

$$\beta_k = \frac{\sigma_{mk}}{\sigma_m^2}.$$  \hfill (2.68)

It says that the only individual thing influencing an asset’s risk is its covariance with the market portfolio. To see why this makes sense, let us reexamine the variance of a portfolio with $N$ assets. It is

$$\sigma_p^2 = \sum_{j=1}^{N} \sigma_j^2 + \sum_{j=1}^{N} \sum_{k=1, k \neq j}^{N} \sigma_{jk}. $$ \hfill (2.69)

For the sake of simplicity, we will assume that all $N$ assets are equally weighted. Thus, we can write
\[
\sigma_p^2 = \sum_{j=1}^{N} \left( \frac{1}{N} \right)^2 \sigma_j^2 + \sum_{j=1}^{N} \sum_{k=1 \atop k \neq j}^{N} \left( \frac{1}{N} \right) \sigma_{jk} \\
= \left( \frac{1}{N} \right) \sum_{j=1}^{N} \sigma_j^2 + \left( \frac{N-1}{N} \right) \sum_{j=1}^{N} \sum_{k=1 \atop k \neq j}^{N} \frac{\sigma_{jk}}{N(N-1)}.
\]
(2.70)

It can easily be seen that the first term in brackets is a simple average of the standard deviations. To see that the second term in brackets is also just a simple average, note that there must be \(N(N-1)\) covariances as the second summation sign excludes \(k = j\). Simplifying we get (Elton and Gruber, 1995)

\[
\sigma_p^2 = \left( \frac{1}{N} \right) \bar{\sigma}_j^2 + \left( \frac{N-1}{N} \right) \bar{\sigma}_{jk}.
\]
(2.71)

It follows that the individual variances of the risky assets can be diversified away by increasing the number of different items included in the portfolio. However, there is a part of the portfolio risk that cannot be diversified away. This is average covariance of the included assets. Thus, the covariance is the only relevant term for assessing an asset’s contribution to the total portfolio risk.

**Result 2.5 b:** An investor can theoretically avoid all risk but the part regarding swings in economic activity.

Even efficient portfolios are exposed to covariance risk. Then, it makes sense that the CAPM only rewards investors for their portfolio’s responsiveness to the market. The risk that can be diversified away is often called idiosyncratic or specific risk. The part that cannot be escaped is often referred to as systematic risk or market risk. Beta is a measure of this risk. The intuition discussed here carries through to the SML. Assets that are unconcerned with the ups and downs of the overall market only deliver the risk-free rate. Though, the assets that are highly influenced by economic activity must consequently suggest a higher reward. This is consistent with traditional economic theory as investors normally demand a higher return from the aggressive stocks than from the defensive ones (Sharpe, 1964). A notable drawback to beta as the solemn risk measure is of course that it relies on MVA to hold. If investors care about more than expected return and variance, for example the probability of extreme events, the foundation on which our analysis is based falls apart.
Chapter 3 Econometric techniques for testing the CAPM

In this chapter we develop the statistical framework for testing the CAPM. We will build test statistics and provide the formal derivations for a thorough understanding of the material. First, a few preliminary comments will be made.

For the purpose of statistical testing it is convenient to express the Sharpe-Lintner model in terms of excess returns. That is, in terms of the risk premium you receive in addition to the risk-free rate. Let $Z_i$ and $Z_m$ denote the expected excess return of asset $i$ and the market portfolio respectively. Then, the traditional CAPM model can be written as

$$E[Z_i] = \beta_{im} E[Z_m],$$

(3.1)

where

$$Z_i \equiv R_i - R_f,$$

(3.2)

and

$$\beta_{im} = \frac{Cov[Z_i, Z_m]}{Var[Z_m]},$$

(3.3)

One should note that (3.3) is equivalent to the definition of beta in chapter 2 because the CAPM treats the risk-free rate as a constant. However, when conducting empirical tests, the proxies for the risk-free rate are stochastic. This means that (3.3) can deviate from the beta defined in chapter 2 when estimating the model. The empirical research that tests the CAPM primarily concentrates on three aspects of (3.1):

1. There is no intercept.
2. Beta can explain the variation in the cross-sections of expected excess returns perfectly.
3. The risk premium in the market is positive.

We will focus on the first implication. After a thorough treatment of this, we turn our attention to the second and third implication. Before discussing the intuition behind testing the first implication, the testable market model is introduced.

3.1 The Market Model

In the traditional Sharpe-Lintner version, as well as the excess return model in (3.1), there is no time dimension as the CAPM is a one-period model. However, when conducting empirical tests of the model, we use panel data in order to apply the econometric methods. Thus, there is a time dimension in the sample data. Therefore, asset returns are assumed to be independently and identically distributed over time (IID) and jointly multivariate normal. Then, it makes sense to make single estimates of the model parameters using data collected over time. Intuitively, when the properties of the data do not change over time, the CAPM can theoretically hold period by period. Additionally, as discussed in chapter 2, the normality assumption is consistent with mean variance analysis because higher order moments can be expressed as functions of the first and second moment in the normal distribution. The assumptions that returns are IID and jointly normal are strong, and there is vast empirical evidence indicating that they are too strong. Therefore, we will also derive a test robust to this notion. This can serve as a check that the inferences are not biased by the distributional assumptions. In our analysis $N$ will denote the number of assets or portfolios under consideration, and $T$ will denote the number of periods under consideration. Consider the market model
Result 3.1: If one or more intercepts in the vector $\mathbf{\alpha}$ are statistically significantly different from zero, we infer, at a certain significance level, that the CAPM does not hold. If the converse is true, we cannot reject the CAPM at that significance level.

\[ Z_t = \mathbf{\alpha} + \mathbf{\beta} Z_{mt} + \mathbf{\epsilon}_t \]  
(3.4)

where

\[ E[\mathbf{\epsilon}_t] = 0, \]  
(3.5)

\[ E[\mathbf{\epsilon}_t \mathbf{\epsilon}_t^\top] = \mathbf{\Sigma}, \]  
(3.6)

and

\[ Cov[Z_{mt}, \mathbf{\epsilon}_t] = 0. \]  
(3.7)

Here $Z_t$ denotes the $N \times 1$-vector of expected excess returns at time $t$, $\mathbf{\alpha}$ is the $N \times 1$-vector of intercepts, $\mathbf{\beta}$ the $N \times 1$-vector of betas, the scalar $Z_{mt}$ is the market risk premium at time $t$, and $\mathbf{\epsilon}_t$ is the $N \times 1$-vector of error terms. Note that the notion of temporal IID returns does not imply anything about the cross-sections of data. In fact, there should be some correlation between returns for the CAPM, and in particular beta, to make sense. Thus, (3.6) implies that there can be heteroskedasticity (different variances in the diagonal of the matrix) and cross-correlation (other terms in the matrix above zero) in the cross-sections of data. Accordingly, the $N \times N$-matrix $\mathbf{\Sigma}$ contains the variances and covariances of the disturbances. From the above assumptions regarding the behavior of returns over time, we can say the following about the vector of error terms

\[ \mathbf{\epsilon}_t \sim N(0, \mathbf{\Sigma}). \]

That is, the error terms follow a multivariate normal distribution with a mean vector of zeros $\mathbf{0}$ and a variance-covariance-matrix $\mathbf{\Sigma}$ (Campbell et al., 1997). The assumption stated in (3.7) says that there is no endogeneity in the model. Endogeneity can result from omitting a relevant variable correlated with the explaining variable already in the model and can cause estimation biases (Wooldridge, 2009).

When comparing (3.4) with (3.1) it is obvious that if the CAPM holds, the intercepts in the vector $\mathbf{\alpha}$ should all be zero. To see why this must apply, note that the traditional CAPM implies an exact linear relationship between beta and expected return. Under the assumptions listed in chapter 2, the third derivation of the CAPM explicitly showed that this relationship applies, by mathematical fact, only for efficient portfolios. Thus, if there is not an exact linear relationship between beta and expected returns in the estimated model, the market portfolio (or its proxy) cannot be mean variance efficient. It should be pointed out, that the notion of “not exactly linear” does not necessarily mean that there is a non-linear relationship. It can also be seen as the restriction that there must be no dispersion around the line relating the two variables (the SML). This implies that beta perfectly predicts expected returns. Testing whether the elements in $\mathbf{\alpha}$ equals zero can be seen as a test of the exact linear relationship expressed in the CAPM. If this relationship does not exist, the market portfolio is not efficient and the CAPM is rejected. Also, if the market portfolio is not mean variance efficient, it cannot be the tangency portfolio. Hence, testing whether the intercepts are zero can also be interpreted as testing whether the market portfolio is the tangency portfolio. The intercepts can also be understood as the Jensen’s alphas from performance evaluation mentioned in section 2.5. When the intercepts are all zero, every asset must lie on the SML just as the CAPM predicts (Roll, 1977). This gives the following result:

Result 3.1: If one or more intercepts in the vector $\mathbf{\alpha}$ are statistically significantly different from zero, we infer, at a certain significance level, that the CAPM does not hold. If the converse is true, we cannot reject the CAPM at that significance level.
Using result 3.1 we can set up the main hypothesis of this chapter

\[ H_0: \alpha = 0 \]  \hspace{1cm} (3.8)

against

\[ H_A: \alpha \neq 0. \]  \hspace{1cm} (3.9)

Notice that this is a \textit{joint} hypothesis as we simultaneously test if every intercept \( \alpha_i \) is zero. To test the \( H_0 \) against the alternative, we will use several test statistics. To implement these tests, we first have to estimate the parameters of the model \( \alpha, \beta \) and \( \Sigma \). Here we will apply Maximum Likelihood Estimation (MLE). As the single independent variable \( Z_{mt} \) is the same in all the equations expressed by the vectors of (3.4), the OLS-estimators of \( \alpha \) and \( \beta \) are identical to the ML-estimators (Cuthbertson and Nitzsche, 2004). Generally, it can be shown that the OLS-estimators are equivalent to ML-estimators when the Gauss-Markov assumptions are met, and the error term is normally distributed as is the case here (Eliason, 1993). First, the estimators will be derived through the ML approach and then by using OLS.

Under the assumption that returns are IID and jointly multivariate normal, we can define the joint probability density function (pdf) of excess returns conditional of the market risk premium. Given this pdf, we can then estimate the parameters using MLE. For the general case \( X \sim N(\mu, \Sigma) \), the multivariate normal pdf can be written as

\[
f_X(x) = \frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} \times \exp \left[ -\frac{1}{2} (x - \mu)' \Sigma^{-1} (x - \mu) \right],
\]

(3.10)

where \( |\Sigma| \) denotes the determinant of the variance term \( \Sigma \) (Skovmand, 2011 b). This is a multidimensional generalization of the well-known bell-curve. In our regression model (3.4) the variance of the dependent variable \( Z_t \) conditional on the independent variable \( Z_{mt} \) is by definition equal to the variance of the error term (3.6). Furthermore, for the mean-vector \( \mu \) in (3.10), the error term obviously disappears when computing the conditional expected value of the dependent variable in the regression model (Wooldridge, 2009). Using (3.10) we have the following multivariate pdf for the jointly normal excess returns in time \( t \) conditional on the observed market risk premium (Campbell et al. 1997)

\[
f(Z_t | Z_{mt}) = \frac{1}{(2\pi)^{N/2} |\Sigma|^{1/2}} \times \exp \left[ -\frac{1}{2} (Z_t - (\alpha + \beta Z_{mt}))' \Sigma^{-1} (Z_t - (\alpha + \beta Z_{mt})) \right]
\]

(3.11)

In order to express the \textit{joint} pdf for every period, we use the fact that returns are assumed to be IID through time. The joint pdf of independent random variables is merely the product of them (Eliason, 1993). Thus, using (3.11) the conditional joint pdf of the excess return vectors \( Z_1, Z_2, \ldots, Z_T \) is
\[
f(Z_1, Z_2, \ldots, Z_T | Z_{m1}, Z_{m2}, \ldots, Z_{mT}) = \prod_{t=1}^{T} f(Z_t | Z_{mt}), \tag{3.12}
\]

which states that the joint pdf is the product of the sequence of pdf’s for the single periods. Substituting (3.11) into (3.12) we get
\[
f(Z_1, Z_2, \ldots, Z_T | Z_{m1}, Z_{m2}, \ldots, Z_{mT}) = \prod_{t=1}^{T} \frac{1}{N^T} \frac{1}{(2\pi)^{T/2} |\Sigma|^{1/2}} \times \exp \left[-\frac{1}{2} (Z_t - \alpha - \beta Z_{mt})' \Sigma^{-1} (Z_t - \alpha - \beta Z_{mt}) \right]. \tag{3.13}
\]

Built on the assumptions that returns are jointly multivariate normal and temporal independent, we now have an expression for the joint pdf of excess returns conditional on the market excess return. As we have determined which distribution the excess returns follow, we can use ML to estimate the market model (3.4) given the observations. We apply this method as the MLE estimators have very attractive properties when what is known as the regularity conditions are met. These conditions will not be discussed here, and we will assume that they are fulfilled. Thus, the ML estimators we derive are consistent, asymptotically efficient, and asymptotically normal (Campbell et al., 1997).

To understand the method of MLE, let us first review the concept of a probability density function. A pdf is a function assigning a probability to each possible value or range of values of a continuous random variable. Consider a sample of observations of the variables in the market model. MLE finds estimates for the unknown parameters of the market model \( \alpha, \beta, \) and \( \Sigma \) that maximizes the likelihood of observing exactly those values of the variables. This is equivalent to saying that we find estimates that maximize the joint pdf (3.13) given the sample of data (Eliason, 1993). In order to find the maximum likelihood estimators we form the log-likelihood function \( \mathcal{L} \) where the parameters of the market model are variables. This is done by taking the log of (3.13)\(^{11}\)

\[
\mathcal{L}(\alpha, \beta, \Sigma) = \sum_{t=1}^{T} \log \left[ \frac{1}{N^T (2\pi)^{T/2} |\Sigma|^{1/2}} \right] - \frac{1}{2} \sum_{t=1}^{T} (Z_t - \alpha - \beta Z_{mt})' \Sigma^{-1} (Z_t - \alpha - \beta Z_{mt})
= -T \left[ \log \left[ \frac{N}{(2\pi)^T} \right] + \log \left[ |\Sigma|^{1/2} \right] \right] - \frac{1}{2} \sum_{t=1}^{T} (Z_t - \alpha - \beta Z_{mt})' \Sigma^{-1} (Z_t - \alpha - \beta Z_{mt})
= -\frac{NT}{2} \log(2\pi) - \frac{T}{2} \log |\Sigma| - \frac{1}{2} \sum_{t=1}^{T} (Z_t - \alpha - \beta Z_{mt})' \Sigma^{-1} (Z_t - \alpha - \beta Z_{mt}). \tag{3.14}
\]

The first equality follows from the rule that taking the log of the product of a sequence reduces the expression to the sum of a sequence of logs. The next equalities follow from the basic rules of log terms. Consistent with the method described above, we find the values of \( \alpha, \beta, \) and \( \Sigma \) that maximize the log-likelihood function. This is a basic maximization problem as solved in the second

\(^{11}\) Working with sums is a lot simpler than working with products.
derivation of the CAPM. Accordingly, we differentiate (3.14) with respect to each of the parameters and set these partial derivatives equal to zero.

To get the first partial derivative we use the rule \( \partial[(x - s)'W(x - s)]/\partial s = -2W(x - s) \), where \( W \) is a symmetric matrix. As the variance-covariance matrix \( \Sigma \) is symmetric, the rule can be applied by letting \( W = \Sigma^{-1} \), \( s = \alpha \), and \( x = Z_t - \beta Z_{mt} \). We have

\[
\frac{\partial L}{\partial \alpha} = -\frac{1}{2} \sum_{t=1}^{T} -2\Sigma^{-1}(Z_t - \alpha - \beta Z_{mt})
= \Sigma^{-1} \left[ \sum_{t=1}^{T} (Z_t - \alpha - \beta Z_{mt}) \right].
\]

(3.15)

For the next partial derivative let \( A = \beta \), \( s = Z_{mt} \), and \( x = Z_t - \alpha \). Also, note that \( Z_{mt} \) is a scalar and that the transpose of a scalar is the same scalar. The first equality follows from the rule \( \partial[(x - As)'W(x - As)]/\partial A = -2W(x - As)s' \), where \( W \) is a symmetric matrix. We get

\[
\frac{\partial L}{\partial \beta} = -\frac{1}{2} \sum_{t=1}^{T} -2\Sigma^{-1}(Z_t - \alpha - \beta Z_{mt})Z_{mt}
= \Sigma^{-1} \left[ \sum_{t=1}^{T} (Z_t - \alpha - \beta Z_{mt}) Z_{mt} \right].
\]

(3.16)

Now, let \( X = \Sigma \) and \( a = b = (Z_t - \alpha - \beta Z_{mt}) \). First, we use the rule \( \partial \ln(|X|)/\partial X = (X')^{-1} \) and recall that \( \Sigma' = \Sigma \) as it is symmetric. Then, we utilize that \( \partial (a'X^{-1}b)/\partial X = -[X']^{-1}ab'[X']^{-1} \) also noting that \( \Sigma \) is symmetric. We write

\[
\frac{\partial L}{\partial \Sigma} = -\frac{T}{2} \Sigma^{-1} - \frac{1}{2} \sum_{t=1}^{T} -\Sigma^{-1}(Z_t - \alpha - \beta Z_{mt})(Z_t - \alpha - \beta Z_{mt})'\Sigma^{-1}
= -\frac{T}{2} \Sigma^{-1} + \frac{1}{2} \Sigma^{-1} \left[ \sum_{t=1}^{T} (Z_t - \alpha - \beta Z_{mt})(Z_t - \alpha - \beta Z_{mt})' \right] \Sigma^{-1}.
\]

(3.17)

For the matrix rules applied here see Petersen and Pedersen (2008). Setting the partial derivatives in (3.15), (3.16), and (3.17) equal to zero, we can then solve for the maximum likelihood estimators. First we solve for the estimator of the intercept vector \( \hat{\alpha} \)

\[
\leftrightarrow \hat{\Sigma}^{-1} \left[ \sum_{t=1}^{T} (Z_t - \hat{\alpha} - \hat{\beta} Z_{mt}) \right] = 0
\]

\[
\leftrightarrow \sum_{t=1}^{T} (Z_t - \hat{\beta} Z_{mt}) = \hat{\alpha} T
\]

\[
\leftrightarrow \hat{\alpha} = \hat{\mu} - \hat{\beta} \hat{\mu}_m
\]

(3.18)
where
\[ \hat{\mu} = \frac{1}{T} \sum_{t=1}^{T} (Z_t) \] (3.19)
and
\[ \hat{\mu}_m = \frac{1}{T} \sum_{t=1}^{T} (Z_{mt}). \] (3.20)

The second equality follows from \( \hat{\alpha} \) having no time dimension. Setting (3.16) equal to zero we get
\[ \mathbf{\Sigma}^{-1} \left[ \sum_{t=1}^{T} (Z_t - \hat{\alpha} - \hat{\beta} Z_{mt}) Z_{mt} \right] = 0 \]
\[ \iff \sum_{t=1}^{T} (Z_t - (\hat{\mu} - \hat{\beta} \hat{\mu}_m) - \hat{\beta} Z_{mt}) = 0 \]
\[ \iff \sum_{t=1}^{T} (Z_t - \hat{\mu}) = - \sum_{t=1}^{T} \hat{\beta} (\hat{\mu}_m - Z_{mt}) \]
\[ \iff \hat{\beta} = \frac{\sum_{t=1}^{T} (Z_t - \hat{\mu}) (Z_{mt} - \hat{\mu}_m)}{\sum_{t=1}^{T} (Z_{mt} - \hat{\mu}_m)^2}. \] (3.21)

The second equality comes from inserting (3.18) and the fourth equality from rearranging and multiplying the numerator and denominator by \((Z_{mt} - \hat{\mu}_m)\). Note that (3.21) looks a lot like the former definitions of beta with the covariance between the assets and the market in the numerator and the variance of the market portfolio in the denominator. Finally, we solve for the estimator in (3.17). We write
\[ \iff - \frac{T}{2} \mathbf{\Sigma}^{-1} + \frac{1}{2} \mathbf{\Sigma}^{-1} \left[ \sum_{t=1}^{T} (Z_t - \hat{\alpha} - \hat{\beta} Z_{mt}) (Z_t - \hat{\alpha} - \hat{\beta} Z_{mt})' \right] \mathbf{\Sigma}^{-1} = 0 \]
\[ \iff \frac{1}{2} \sum_{t=1}^{T} (Z_t - \hat{\alpha} - \hat{\beta} Z_{mt}) (Z_t - \hat{\alpha} - \hat{\beta} Z_{mt})' \mathbf{\Sigma}^{-1} = \frac{T}{2} \]
\[ \iff \mathbf{\Phi} = \frac{1}{T} \sum_{t=1}^{T} (Z_t - \hat{\alpha} - \hat{\beta} Z_{mt}) (Z_t - \hat{\alpha} - \hat{\beta} Z_{mt})'. \] (3.22)

The second equality comes from multiplying each side by \( \mathbf{\Phi}^{-1} \) (Campbell et al., 1997). It should be emphasized that in order to find the optimal parameters all the three partial derivatives must equal zero. In the derivations above we solved the equations isolated one by one. Note however, that the solutions (3.18) and (3.21) are explicitly present in the third solution (3.22). Thus, the estimator \( \mathbf{\Phi} \) satisfies the criteria that all the partial derivatives equal zero. In the derivation of (3.21) we inserted (3.18), so two of the partial derivatives must equal zero for the solution of \( \hat{\beta} \). Furthermore, if we

\[ \text{As its inverse is present in each term, we get the identity matrix } I. \text{ We then use the rule that } I_n A = A, \text{ where } A \text{ is a } k \times n-\text{matrix. In principle, this trick was also used in the derivations of (3.18) and (3.21). Alternative one could argue that we only had to solve the term in the brackets in those cases.} \]
inserted the inverse of (3.22) in the first line of the derivation of (3.21), the solution would obviously be the same. A similar logic applies to the first estimator \( \hat{\alpha} \), as (3.21) is explicitly included in the result and (3.22) would cancel out in the derivation.

Earlier we claimed that the ML estimators (3.18), (3.21), and (3.22) are equivalent to the OLS estimators. We will now prove that by deriving the OLS estimators for each asset and argue that they are equivalent to the expressions above. To do that we apply the method of least squares. Again, the overall idea is to find the parameter values that fit the data the best. One way to solve this problem is by minimizing the sum of squared residuals (SSR). In each period the market model equation for asset \( i \) is

\[
Z_{it} = \alpha_i + \beta_i Z_{mt} + \epsilon_{it}. \tag{3.23}
\]

Note that all the terms are scalars and the assumptions of temporal IID normal errors and no endogeneity also apply here. The SSR for asset \( i \) is

\[
SSR = \sum_{t=1}^{T} (\epsilon_{it})^2 = \sum_{t=1}^{T} (Z_{it} - (\alpha_i + \beta_i Z_{mt}))^2. \tag{3.24}
\]

Minimizing SSR is a simple quadratic problem, so we take the partial derivatives with respect to the parameters and set them equal to zero (Skovmand, 2011 b). To take the derivative of (3.24) we apply the chain rule. We get

\[
\frac{\partial SSR}{\partial \alpha_i} = \sum_{t=1}^{T} 2(Z_{it} - (\alpha_i + \beta_i Z_{mt})) \tag{3.25}
\]

and

\[
\frac{\partial SSR}{\partial \beta_i} = \sum_{t=1}^{T} -2(Z_{it} - (\alpha_i + \beta_i Z_{mt})) \times Z_{mt}. \tag{3.26}
\]

We then set (3.25) and (3.26) equal to zero and solve for the estimators. First, we write

\[
\sum_{t=1}^{T} 2 \left( Z_{it} - (\hat{\alpha}_i + \hat{\beta}_i Z_{mt}) \right) = 0
\]

\[
\leftrightarrow \sum_{t=1}^{T} Z_{it} - T \hat{\alpha}_i - \hat{\beta}_i \sum_{t=1}^{T} Z_{mt} = 0
\]

\[
\leftrightarrow \hat{\alpha}_i = \hat{\mu}_i - \hat{\beta}_i \hat{\mu}_m, \tag{3.27}
\]

where

\[
\hat{\mu}_i = \frac{1}{T} \sum_{t=1}^{T} Z_{it} \tag{3.28}
\]

and
\[ \hat{\mu}_m = \frac{1}{T} \sum_{t=1}^{T} Z_{mt}. \]  

(3.29)

Also, we have

\[ \sum_{t=1}^{T} -2(Z_{it} - (\alpha_i + \beta_i Z_{mt})) \times Z_{mt} = 0 \]

\[ \sum_{t=1}^{T} \left[ Z_{it} - \left( (\hat{\mu}_i - \hat{\beta}_i \hat{\mu}_m) + \hat{\beta}_i Z_{mt} \right) \right] = 0 \]

\[ \sum_{t=1}^{T} \left[ (Z_{it} - \hat{\mu}_i) - \hat{\beta}_i (Z_{mt} - \hat{\mu}_m) \right] = 0 \]

\[ \iff \hat{\beta}_i = \frac{\sum_{t=1}^{T} (Z_{it} - \hat{\mu}_i)(Z_{mt} - \hat{\mu}_m)}{(Z_{mt} - \hat{\mu}_m)^2}. \]  

(3.30)

Note that (3.30) is explicitly included in (3.27), and that (3.27) is explicitly included in (3.30). Thus both partial derivatives equal zero in each estimator. The OLS-estimators (3.27) and (3.30) can be used to estimate the individual intercepts and betas in the vectors of the market model (3.4). Consequently, we can rewrite these two expressions in vector-form to include the estimates for every asset \( i \) and get (3.18) and (3.21). Hence, we have showed that the ML estimators and the OLS estimators are equivalent. Also, when applying OLS, one could use the estimators \( \hat{\alpha} \) and \( \hat{\beta} \) to estimate the missing \( \hat{\Sigma} \) by writing \( \Sigma \) as a function of the two others using (3.4) and (3.6). The result would obviously be (3.22).

We have now derived the ML estimators, and showed that they are equivalent to the OLS estimators. In order to construct a test statistic for testing our hypothesis, we need to know the distribution of the estimators. We have assumed that excess returns are jointly normal and temporally IID. The estimators’ distributions result from this supposition. Conditional on the market risk premium, \( \hat{\alpha} \) and \( \hat{\beta} \) follow a normal distribution (Campbell et al., 1997). Their expected values equal the true parameter values, so they are unbiased. This follows from the specification of the market model, and in particular the exogeneity of the explanatory variable combined with the IID assumption. Their variances can be found by setting up the inverse of the Fisher information matrix. This can be found by setting up the Hessian matrix containing the second order partial derivatives of the log likelihood function (Eliason, 1993). A more straightforward approach will be outlined here. The full derivation is found in appendix B.1. First, we use that \( \text{Cov}(\hat{\mu}, \hat{\beta} \hat{\mu}_m) = 0 \). This result will not be proven here. It should be underlined that \( \hat{\mu} \) is the vector of mean excess returns - not excess returns. Using (3.18) and the independency just stated, we write

\[ \text{Var}(\hat{\alpha}) = \text{Var}(\hat{\mu} - \hat{\beta} \hat{\mu}_m) = \text{Var}(\hat{\mu}) + \text{Var}(\hat{\beta} \hat{\mu}_m). \]  

(3.31)

Substituting (3.19) and (3.21) into this expression we get (Skovmand, 2011 b)

\[ \text{Var}(\hat{\alpha}) = \text{Var} \left[ \frac{1}{T} \sum_{t=1}^{T} (Z_t) \right] + \text{Var} \left[ \frac{\sum_{t=1}^{T} (Z_t - \hat{\mu})(Z_{mt} - \hat{\mu}_m)}{\sum_{t=1}^{T} (Z_{mt} - \hat{\mu}_m)^2} \hat{\mu}_m \right]. \]  

(3.32)
In appendix B.1 we show that this can be reduced to
\[
\text{Var}(\hat{\alpha}) = \frac{1}{T} \left[ 1 + \frac{\hat{\mu}^2_m}{\hat{\sigma}^2_m} \right] \Sigma, \tag{3.33}
\]
where
\[
\hat{\sigma}^2_m = \frac{1}{T} \sum_{t=1}^{T} (Z_{mt} - \hat{\mu}_m)^2. \tag{3.34}
\]
Likewise we argue in appendix B.1 that
\[
\text{Var}(\hat{\beta}) = \frac{1}{T} \left[ \frac{1}{\hat{\sigma}^2_m} \right] \Sigma. \tag{3.35}
\]
Now we only need the properties of the estimator for the variance-covariance matrix of the error terms \( \Sigma \). Here we apply the following theorem (Uhlig, 1994):

Consider the random variable \( Y_t \sim N(0, \Sigma) \) for \( j = 1, 2, \ldots, T \), where \( \Sigma \) is a \( N \times N \)-positive definite matrix. Then
\[
X = \sum_{t=1}^{T} Y_t Y_t' \sim W_N(T, \Sigma). \tag{3.36}
\]
This means that \( X \) follows a Wishart distribution with \( T \) degrees of freedom and variance-covariance matrix \( \Sigma \). The Wishart distribution is the chi-square distribution generalized to a multivariate case. Previously, we found out that \( \epsilon_t \sim N(0, \Sigma) \). Let \( Y_t = \epsilon_t \) and recall from (3.6) that \( E[\epsilon_t \epsilon_t'] = \Sigma \). Then, it can be shown that \( T\Sigma \) follows a Wishart distribution with \( T - 2 \) degrees of freedom and variance-covariance matrix \( \Sigma \).

We can summarize the above conclusions into the following result:

**Result 3.2:**
\[
\bar{\alpha} \sim N \left( \alpha, \frac{1}{T} \left[ 1 + \frac{\hat{\mu}^2_m}{\hat{\sigma}^2_m} \right] \Sigma \right) \tag{3.37}
\]
\[
\bar{\beta} \sim N \left( \beta, \frac{1}{T} \left[ \frac{1}{\hat{\sigma}^2_m} \right] \Sigma \right) \tag{3.38}
\]
\[
T \Sigma \sim W_N(T - 2, \Sigma). \tag{3.39}
\]

### 3.2 The standard tests

Now we have built the necessary foundation to construct test statistics for testing the null hypothesis. When conducting such tests, it is relevant to distinguish between the *unrestricted* and the *restricted* model. The unrestricted model is the model where the null hypothesis is not imposed. This is the market model (3.4). Conversely, in the restricted model the constraints of the null hypothesis are imposed. Recall that the null hypothesis is a *joint* hypothesis. We wish to test simultaneously whether *all* the intercepts are zero. Thus, there are in fact multiple hypotheses. Here, the hypotheses can be seen as exclusion restrictions, as they practically exclude the intercepts from the market model. There are different ways of testing multiple restrictions. One possibility is the
Wald test. This test only requires the estimators of the unrestricted model. These are the estimators we have derived above (Wooldridge, 2009).

In a Wald test, we evaluate the difference between the estimate of the market model parameter and its value under the null hypothesis. Formally, the squared difference is benchmarked against the variance of the parameter. This result is then compared to the test statistic’s distribution under the null hypothesis to see whether the above difference is significant. Using result (3.1) we formed the null hypothesis
\[
\hat{\alpha}^2 = 0.
\]
Thus, the Wald test statistic is
\[
J_0 = \hat{\alpha}' [\text{Var}(\hat{\alpha})]^{-1} \hat{\alpha}.
\]

Under the null hypothesis $J_0$ has a chi-square distribution with $N$ degrees of freedom. Here, $N$ corresponds to the number of restrictions imposed as there are $N$ different intercepts. The problem with (3.40) is that it uses the true value of the parameter $\hat{\Sigma}$ which, of course, is not known. Thus, we need a consistent estimator of this parameter to really use (3.40). Here we can apply the maximum likelihood estimator (3.22). However, when using the estimator the Wald test statistic only has a chi-square distribution with $N$ degrees of freedom asymptotically. In this context, asymptotically should be interpreted as, “the number of periods in the sample goes to infinity”. We can summarize into the following result:

**Result 3.3:** The Wall test of the null hypothesis (3.8) can be defined as
\[
J_0 = T \left[ 1 + \frac{\hat{\mu}_m^2}{\hat{\sigma}_m^2} \right]^{-1} \hat{\alpha}' \hat{\Sigma}^{-1} \hat{\alpha} \sim \chi^2_N.
\] (3.41)

Interestingly, using the above results, we can also construct a test statistic in a finite-sample setting. To do this, we employ the following theorem:

Let $x$ be a $m \times 1$-vector and $A$ be a $m \times m$-matrix. Furthermore, let $x$ and $A$ be independent, $x \sim N(0, \Omega)$, and $A \sim W_N(n, \Omega)$, where $m \leq n$. Then
\[
\frac{(n - m + 1)}{m} x'A^{-1}x \sim F_{m,n-m+1}.
\] (3.42)

First, note that the estimator $\hat{\alpha}$ of the intercept-vector is independent of the estimator $\hat{\Sigma}$ of the variance-covariance matrix. Thus, we can transform (3.41) into a new test statistic with an exact distribution. To do this, we have to fit the elements of (3.41) into (3.42) while meeting the criteria of the theorem. Let
\[
x = \left( T \left[ 1 + \frac{\hat{\mu}_m^2}{\hat{\sigma}_m^2} \right]^{-1} \right)^{1/2} \hat{\alpha}, A = T\hat{\Sigma}, m = N, \text{ and } n = (T - 2).
\]

Note that the only non-scalar in $x$ is $\hat{\alpha}$. According to result 3.2, $x$ and $A$ will, under the null hypothesis $\alpha = 0$, thus have the distributions defined in the above theorem. The criteria that
\( m = N \leq n = T - 2 \) merely relates to how the sample should be drawn. As we will see later, it will probably almost always be satisfied in practice. Inserting into (3.42) we get a new test statistic

\[
J_1 = \frac{(T - 2 - N + 1)}{N}(T \left[ 1 + \frac{\hat{\mu}_m^2}{\hat{\sigma}_m^2} \right]^{-1})^{1/2} \hat{\alpha} \left( T \Sigma^{-1} \right)^{-1/2} \left( T \left[ 1 + \frac{\hat{\mu}_m^2}{\hat{\sigma}_m^2} \right]^{-1} \right)^{1/2} \hat{\alpha}
\]

\[
= \frac{(T - N - 1)}{N} \left( \frac{1 + \frac{\hat{\mu}_m^2}{\hat{\sigma}_m^2}}{\hat{\mu}_m^2} \right)^{-1/2} \left( 1 + \frac{\hat{\mu}_m^2}{\hat{\sigma}_m^2} \right)^{-1/2} T^{1/2} T^{-1/2} \hat{\alpha}' \Sigma^{-1} \hat{\alpha}.
\]

Under the null hypothesis \( J_1 \) has an exact central \( F \) distribution with \( N \) degrees of freedom in the numerator and \( (T - N - 1) \) degrees of freedom in the denominator (Campbell et al., 1997). Recalling (3.40) we reach the next result:

**Result 3.4:** The central \( F \) test of the null hypothesis (3.8) can be defined as

\[
J_1 = \frac{(T - N - 1)}{TN} J_0 \sim F_{N,T-N-1}.
\]  

Now, we have considered two ways of testing the null hypotheses jointly: the Wald test and the central \( F \) test. The second was a transformation of the first. A third way to test multiple restrictions is the likelihood ratio (LR) test. This test measures the difference between the log-likelihood functions of the unrestricted and the restricted model. Recall that the ML estimators maximize the log-likelihood function. Thus, when some parameters are restricted to zero, the log-likelihood must decrease (or stay unchanged). Applying the LR test, we measure whether this drop is significant meaning that the excluded variables are influential (Wooldridge, 2009). In terms of our analysis, a significant drop in the log-likelihood would contradict the efficiency of the market portfolio. To apply the LR test, we obviously need the estimators of the restricted model. Setting \( \alpha \) to zero in (3.14), we get the log-likelihood function of the constrained model \( \mathcal{L}^* \). We write

\[
\mathcal{L}^*(\beta, \Sigma) = -\frac{NT}{2} \log(2\pi) - \frac{T}{2} \log|\Sigma| - \frac{1}{2} \sum_{t=1}^{T} (Z_t - \beta Z_{mt})' \Sigma^{-1} (Z_t - \beta Z_{mt}).
\]  

The same approach we took to derive (3.21) and (3.22) can be used. Thus (3.46) is differentiated partially with respect to the remaining parameters. Then, the results are set equal to zero and solved simultaneously to derive the expressions for the parameter estimators of the constrained model. Here we can skip some algebra by setting \( \alpha \) to zero in (3.16) and (3.17). The estimators are

\[
\hat{\beta}^* = \frac{\sum_{t=1}^{T} Z_t Z_{mt}}{\sum_{t=1}^{T} Z_{mt}^2}
\]  

and

38
Using similar reasoning as outlined under the unconstrained model, and in appendix B.1, the distributions of the constrained estimators can also be derived. They are (Campbell et al., 1997)

$$\tilde{\beta}^* \sim N \left( \beta, \frac{1}{T} \left[ \frac{1}{\hat{\mu}_m^2 + \hat{\sigma}_m^2} \right] \Sigma \right), \quad (3.48)$$

and

$$T \tilde{\Sigma}^* \sim W_N \left( T - 1, \Sigma \right). \quad (3.49)$$

To see the link between (3.46) and the variance in (3.48) note, in excess of the derivation in appendix B.1, that

$$T(\hat{\mu}_m^2 + \hat{\sigma}_m^2) = T \left[ \hat{\mu}_m^2 + \frac{1}{T} \sum_{t=1}^{T} (Z_{mt} - \hat{\mu}_m)^2 \right]$$

$$= T \left[ \hat{\mu}_m^2 + \frac{1}{T} \sum_{t=1}^{T} (Z_{mt}^2 + \hat{\mu}_m^2 - 2Z_{mt}\hat{\mu}_m) \right]$$

$$= \sum_{t=1}^{T} Z_{mt}^2. \quad (3.50)$$

In order to construct the test statistic, we first have to define the log-likelihood ratio. The conditional joint pdf of excess returns (3.13) is the likelihood function of the unconstrained model. Imposing the null hypothesis we get the likelihood function of the constrained model. The ratio of the two evaluated at the ML estimators is called the likelihood ratio (Eliason, 1993). Taking logs, we get the log-likelihood ratio. This equals the difference between the log-likelihood functions of the constrained and the unconstrained model also evaluated at the ML estimators. We write

$$LR = L^* - L. \quad (3.51)$$

The expressions of the log-likelihood functions are rather voluminous. Fortunately, they can be reduced considerably. We will now show that the sum-term in the functions can be rewritten into a much simpler form. First, recall that $$(Z_t - \alpha - \beta Z_{mt})$$ is a $N \times 1$-vector and that $\Sigma$ is a $N \times N$-matrix. Thus, the product in the sum-term is actually a scalar and the trace of a scalar is equal to the same scalar. For the unconstrained function, we have

$$\tilde{\Sigma}^* = \frac{1}{T} \sum_{t=1}^{T} (Z_t - \tilde{\beta}^* Z_{mt})(Z_t - \tilde{\beta}^* Z_{mt})'.$$  \quad (3.47)
\[
\sum_{t=1}^{T} (Z_t - \hat{\alpha} - \hat{\beta}Z_{mt})' \Sigma^{-1}(Z_t - \hat{\alpha} - \hat{\beta}Z_{mt})
\]

\[
= \sum_{t=1}^{T} \text{trace} \left[ (Z_t - \hat{\alpha} - \hat{\beta}Z_{mt})' \Sigma^{-1}(Z_t - \hat{\alpha} - \hat{\beta}Z_{mt}) \right]
\]

\[
= \sum_{t=1}^{T} \text{trace} \left[ \Sigma^{-1}(Z_t - \hat{\alpha} - \hat{\beta}Z_{mt})(Z_t - \hat{\alpha} - \hat{\beta}Z_{mt})' \right]
\]

\[
= \text{trace} \left[ \Sigma^{-1} \sum_{t=1}^{T} (Z_t - \hat{\alpha} - \hat{\beta}Z_{mt})(Z_t - \hat{\alpha} - \hat{\beta}Z_{mt})' \right]
\]

\[
= \text{trace} \left[ \Sigma^{-1} T \Sigma \right] = T \text{trace} [I] = NT. \quad (3.52)
\]

Let \( A = (Z_t - \alpha - \beta Z_{mt})' \) and \( B = \Sigma^{-1}(Z_t - \alpha - \beta Z_{mt}) \). The second equality comes from the rule that \( \text{trace}(AB) = \text{trace}(BA) \). The third equality utilizes that the trace of a sum is the same as the sum of a trace. The fourth equality follows from recalling the expression for the estimator (3.22). The last equality is a result of the identity matrix having \( N \) ones in the diagonal. The equivalent computations for the constrained model are very similar, only that \( \hat{\alpha} \) would not appear, and that we would substitute for the constrained estimator (3.47). The result is obviously the same as (3.52). Using (3.52) the log-likelihood ratio (3.51) can be written as

\[
\mathcal{L}R = \left[ -\frac{NT}{2} \log(2\pi) - \frac{T}{2} \log|\Sigma^*| - \frac{1}{2} NT \right] - \left[ -\frac{NT}{2} \log(2\pi) - \frac{T}{2} \log|\hat{\Sigma}| - \frac{1}{2} NT \right]
\]

\[
= -\frac{T}{2} \log|\Sigma^*| + \frac{T}{2} \log|\Sigma|
\]

\[
= -\frac{T}{2} \left[ \log|\Sigma^*| - \log|\Sigma| \right]. \quad (3.53)
\]

The LR test statistic is the log-likelihood ratio (3.51) multiplied by minus 2. We have

\[
J_2 = -2\mathcal{L}R \quad = T[\log|\Sigma^*| - \log|\Sigma|]. \quad (3.54)
\]

If the null hypothesis is true, the LR test has a chi-square distribution asymptotically. The degrees of freedom is the number of excluded variables. This translates into the following result:

**Result 3.5:** The likelihood ratio test of the null hypothesis (3.8) can be defined as

\[
J_2 = (T[\log|\Sigma^*| - \log|\Sigma|])^a \sim \chi^2_N. \quad (3.55)
\]
The conclusions of the asymptotic tests are really only reliable when an infinite number of periods are included in the sample. However, applying a modification to $J_2$ proposed by Jobson and Korkie (1982), we get a test statistic with better properties in finite samples.

**Result 3.6:** The adjusted likelihood ratio test of the null hypothesis (3.8) can be defined as

$$J_3 = \frac{T - \frac{N}{2} - 2}{T} j_2 \sim \chi^2_N.$$ (3.56)

It should be underlined that $J_3$ is still only asymptotically chi-square distributed with $N$ degrees of freedom under the null-hypothesis (Campbell et al., 1997).

We have now derived four different test statistics for the joint test of whether the intercepts in the market model are zero. Interestingly, there exists a mathematical relationship between all four statistics. We will now show the link between the exact test statistic $J_3$ and the large sample statistics. The first relationship was presented in result 3.4. Using a number of algebraic manipulations, Campbell et al. (1997) derives the connection between $J_3$ and $J_2$ arguing that the exact test is actually also a likelihood ratio test. Using (3.56) this connection can be extended to the relationship between $J_1$ and $J_3$:

**Result 3.7:**

$$J_1 = \frac{(T - N - 1)}{TN} J_0 = \frac{(T - N - 1)}{N} \left( \exp \left[ \frac{J_2}{T} \right] - 1 \right).$$ (3.57)

$$J_1 = \frac{(T - N - 1)}{N} \left( \exp \left[ \frac{J_3}{T - \frac{N}{2} - 2} \right] - 1 \right).$$ (3.58)

$$J_1 = \frac{(T - N - 1)}{N} \left( \exp \left[ \frac{J_3}{T - \frac{N}{2} - 2} \right] - 1 \right).$$ (3.59)

Earlier, we argued that the exact linear relationship expressed in the CAPM can only hold if the market portfolio is mean variance efficient. This led us to result 3.1 and the null hypothesis $\alpha = 0$. The economic and mathematical intuition carries through to the test statistics derived so far. In their 1989 paper, Gibbons, Ross, and Shanken derives the result $\tilde{\alpha} \tilde{\Sigma}^{-1} \tilde{\alpha} = \hat{\beta}_q^2 / \hat{\sigma}_q^2 - \hat{\beta}_m^2 / \hat{\sigma}_m^2$. Using (3.43), and noticing that these fractions are in fact just squared Sharpe ratios, leads to an alternative specification of $J_1$:

**Result 3.8:**

$$J_1 = \frac{(T - N - 1)}{N} \left[ \frac{\theta_q^2 - \theta_m^2}{1 + \theta_m^2} \right].$$ (3.60)

where $q$ is the realized tangency portfolio.

Recall from chapter 2 that the tangency portfolio has the highest Sharpe ratio of all risky portfolios. It is the only pure risk portfolio on the SML. All other efficient portfolios are combinations of the risk-free asset and this portfolio. If the market portfolio, consisting of all risky assets, does not have the maximal Sharpe’s ratio, it is not placed on the SML and consequently it cannot be mean variance efficient. This point was also evident in the second derivation in chapter 2 where we maximized the Sharpe ratio to obtain the exact linear relationship of the CAPM. According to (3.60) $J_1$ is a decreasing function of $\theta_m^2$. Thus, it gets easier to reject the efficiency of the market portfolio as the difference between the Sharpe ratios of the tangency and the market portfolio
increases. On the other hand, if the market portfolio equals the tangency portfolio, $J_1$ will be zero, and our null hypothesis cannot be rejected. In result 3.7 we showed that all the statistics derived so far are related functionally. Thus, the intuition just presented also applies for the three other tests $J_0$, $J_2$, and $J_3$.

3.3 Size of the tests
So far we have derived three asymptotic test statistics. Under the null hypothesis $J_0$, $J_2$, and $J_3$ are all asymptotically chi-square distributed with $N$ degrees of freedom. Strictly speaking, this asymptotic result is only accurate in an infinite sample. In empirical applications we must work with finite samples, and they may not be large enough to allow for asymptotic approximations. This introduces a problem regarding the size of the tests. The size of a test is the probability of rejecting $H_0$ when $H_0$ is in fact true. This can be stated as $P(J > c_0 | H_0)$, where $c_0$ is the critical value under $H_0$. When conducting empirical tests, we set a certain significance level and thereby accept that this is the probability of mistakenly rejecting $H_0$. However, when using asymptotic approximations, this probability can be a lot higher than intended.

Fortunately, we have identified a functional relationship between the asymptotic test statistics and the exact $F$-test. We can apply these relationships to calculate how far off the finite-sample sizes of the tests are. Of course, the size of the exact $F$-test always equals the desired significance level as this is already a finite-sample test. To calculate the true size, we first find the critical value under the given asymptotic distribution. Then this value is transformed into the exact $F$ distribution, where the corresponding size can be determined. Consider testing 20 portfolios or securities and 60 months of excess returns with the Wald test $J_0$. We use a 5 % significance level. Under the null hypothesis, $J_0$ has a chi-square distribution with 20 degrees of freedom asymptotically. In this distribution, the critical value for $J_0$ corresponding to a size of 5 % is 31.41. Using (3.57) this value corresponds to a critical value of 1.021 for the exact test $J_1$. Under the null hypothesis, $J_1$ has an exact $F$ distribution with 20 degrees of freedom in the numerator and 39 degrees of freedom in the denominator. In this distribution, the probability of getting a test statistic higher than 1.021 is 46.2 %. This is the true size of $J_0$ in a finite sample of 60 months when considering an asymptotic size of 5 %. The example is illustrated in figure 3.1.

Figure 3.1: Complementary CDFs of an chi-square distribution (left) and a central $F$ distribution (right)

| Source: Author’s creation. |
| Note: The chi-square distribution has 20 degrees of freedom while the central $F$ distribution has 20 degrees of freedom in the numerator and 39 degrees of freedom in the denominator. |

---

13 This commonly referred to as a type I error.
On the vertical axis in the exact $F$ distribution, it is evident that the exact size (the dashed red line) is a lot larger than the intended size (the dashed blue line). The actual threshold of when we reject $H_0$ using the asymptotic $J_0$ test is thus way too strict in a finite sample of 60 months. In a 5 % test, we should reject $H_0$ with $p$-values below 5 %. However, as we use an asymptotic approximation, the limit of 5 % practically works as a limit of 46.2 % in this example. Consequently, we reject 9 times too often.

Table 3.1: True size of the standard asymptotic tests with an asymptotic size of 5 %.

<table>
<thead>
<tr>
<th>$N$</th>
<th>$T$</th>
<th>$J_0$</th>
<th>$J_2$</th>
<th>$J_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>60</td>
<td>0.170</td>
<td>0.096</td>
<td>0.051</td>
</tr>
<tr>
<td></td>
<td>120</td>
<td>0.099</td>
<td>0.070</td>
<td>0.050</td>
</tr>
<tr>
<td></td>
<td>360</td>
<td>0.064</td>
<td>0.056</td>
<td>0.050</td>
</tr>
<tr>
<td></td>
<td>900</td>
<td>0.055</td>
<td>0.052</td>
<td>0.050</td>
</tr>
<tr>
<td>20</td>
<td>60</td>
<td>0.462</td>
<td>0.211</td>
<td>0.057</td>
</tr>
<tr>
<td></td>
<td>120</td>
<td>0.200</td>
<td>0.105</td>
<td>0.051</td>
</tr>
<tr>
<td></td>
<td>360</td>
<td>0.086</td>
<td>0.064</td>
<td>0.050</td>
</tr>
<tr>
<td></td>
<td>900</td>
<td>0.063</td>
<td>0.055</td>
<td>0.050</td>
</tr>
<tr>
<td>50</td>
<td>60</td>
<td>1.000</td>
<td>0.987</td>
<td>0.404</td>
</tr>
<tr>
<td></td>
<td>120</td>
<td>0.826</td>
<td>0.432</td>
<td>0.068</td>
</tr>
<tr>
<td></td>
<td>360</td>
<td>0.228</td>
<td>0.114</td>
<td>0.051</td>
</tr>
<tr>
<td></td>
<td>900</td>
<td>0.101</td>
<td>0.070</td>
<td>0.050</td>
</tr>
</tbody>
</table>

Source: Campbell et al., 1997 and the author’s own calculations.

In table 3.1, the procedure illustrated above is extended to different values of $N$ and $T$ still using an asymptotic size of 5 % and the formulae (3.57)-(3.59). It is apparent that the finite-sample correction of $J_2$ works impressively well. Thus, the exact size of $J_3$ is very close to the desired level in all combinations of $N$ and $T$ but one. Considering the two other asymptotic tests, they generally reject $H_0$ too often. The problem grows as the number of portfolios is increased. On the other hand, the true size converges to the asymptotic size when the number of periods is increased. Consequently, it seems that the problem is most critical when considering a large number of securities or portfolios and a small number of periods. Testing 5 years of monthly data for $N=50$, the finite-sample sizes of $J_0$ and $J_2$ are app. 100 % and 98.7 % respectively. Then, if the $H_0$ is in fact true, one will nearly always reject it wrongly anyway (Campbell et al., 1997).

We can summarize into the following result:

**Result 3.9:** The true sizes of the three asymptotic tests $J_0$, $J_2$, and $J_3$ are closest to their asymptotic sizes when testing data of a small number of portfolios or securities observed over a long time period.
3.4 Power of the tests
We have just illustrated that one might reject the null hypothesis too often when using asymptotic tests. However, even the exact test may lead to unfortunate inferences. Let us consider the reverse problem: we cannot reject even though the alternative hypothesis is true. Here we want to compute the probability of rejecting given that the alternative is true. This is called the power of a test. To compute the power, we first need to specify the critical value under the null hypothesis. Then, the power is the probability that the test statistic exceeds this value when the alternative is true. Formally, this can be stated as \( P(J > c_1 | H_A) \), where \( c_1 \) is the critical value under \( H_0 \).

We now illustrate the potential problem using the exact test as it is easier to work with since we know its finite-sample distribution. Furthermore, the conclusions we reach with it should be representative. Hence, we do not need to make any transformations. We already know that under \( H_0 \), \( J_1 \) is distributed central \( F \) with \( N \) degrees of freedom in the numerator and \( T - N - 1 \) degrees of freedom in the denominator. As the calculation of the power assumes the alternative is true, we also need to specify the distribution of \( J_1 \) under \( H_A \) conditional on the market risk premium. This is the non-central \( F \) distribution, which is a generalization of the \( F \) distribution. We write

\[
J_1 \sim F_{N,T-N-1}(\delta),
\]

where

\[
\delta = T \left[ 1 + \frac{\mu_q^2}{\sigma_q^2} \right]^{-1} \alpha' \Sigma^{-1} \alpha.
\]

The term \( \delta \) is the noncentrality parameter of the \( F \) distribution. Note that when \( \alpha = 0 \) the null hypothesis is true, and (3.61) becomes the ordinary central \( F \) distribution as \( \delta \) is zero. If this is not the case, \( J_1 \) follows the alternative distribution given in (3.61). Thus, the power of \( J_1 \) equals the probability that this distribution assigns to the event that \( J_1 \) exceeds the critical value from \( H_0 \). To specify the distribution of \( J_1 \) under \( H_A \), we need the values of the elements in the parameter \( \delta \). Here we can luckily take a shortcut by using the result from Gibbons, Ross and Shanken that \( \alpha' \Sigma^{-1} \alpha = \mu_q^2 / \sigma_q^2 - \mu_m^2 / \sigma_m^2 \). As \( H_A \) says that \( \alpha \neq 0 \), alternative distributions will imply a difference between the tangency portfolio \( q \) and the market portfolio \( m \). Let us assume that the market portfolio has an annual mean return of 7% and an annual standard deviation of 18%. Furthermore, let us consider the tangency portfolio of three different scenarios. As the tangency portfolio has the maximum Sharpe ratio, we only consider properties superior to those of the market portfolio. If the tangency portfolio equaled the market portfolio, their Sharpe ratios would be equal and \( \delta \) would be zero. Then, we would actually be reporting the size instead of the power, as the null hypothesis would be true. The power of \( J_1 \) at different values of \( N \) and \( T \) is calculated under each of the three scenarios in table 3.2. We use a size of 5%. Again \( N \) is the number of portfolios or securities, and \( T \) is now the number of months included.

---

14 This is commonly known as a type 2 error.
Table 3.2: Power of the exact F-test with a size of 5% given different properties of the tangency portfolio.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>N = 1</th>
<th>5</th>
<th>10</th>
<th>20</th>
<th>50</th>
</tr>
</thead>
<tbody>
<tr>
<td>A: μ_q = 8%, σ_q = 15%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T = 60</td>
<td>0.125</td>
<td>0.078</td>
<td>0.067</td>
<td>0.059</td>
<td>0.052</td>
</tr>
<tr>
<td></td>
<td>0.206</td>
<td>0.113</td>
<td>0.090</td>
<td>0.074</td>
<td>0.061</td>
</tr>
<tr>
<td></td>
<td>0.509</td>
<td>0.284</td>
<td>0.207</td>
<td>0.150</td>
<td>0.102</td>
</tr>
<tr>
<td></td>
<td>0.881</td>
<td>0.667</td>
<td>0.530</td>
<td>0.388</td>
<td>0.236</td>
</tr>
<tr>
<td>B: μ_q = 10%, σ_q = 15%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T = 60</td>
<td>0.220</td>
<td>0.116</td>
<td>0.090</td>
<td>0.072</td>
<td>0.055</td>
</tr>
<tr>
<td></td>
<td>0.393</td>
<td>0.206</td>
<td>0.150</td>
<td>0.110</td>
<td>0.076</td>
</tr>
<tr>
<td></td>
<td>0.836</td>
<td>0.598</td>
<td>0.460</td>
<td>0.328</td>
<td>0.194</td>
</tr>
<tr>
<td></td>
<td>0.996</td>
<td>0.966</td>
<td>0.915</td>
<td>0.809</td>
<td>0.576</td>
</tr>
<tr>
<td>C: μ_q = 12%, σ_q = 15%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T = 60</td>
<td>0.333</td>
<td>0.169</td>
<td>0.122</td>
<td>0.089</td>
<td>0.058</td>
</tr>
<tr>
<td></td>
<td>0.587</td>
<td>0.334</td>
<td>0.238</td>
<td>0.164</td>
<td>0.098</td>
</tr>
<tr>
<td></td>
<td>0.967</td>
<td>0.846</td>
<td>0.728</td>
<td>0.565</td>
<td>0.340</td>
</tr>
<tr>
<td></td>
<td>1.000</td>
<td>0.999</td>
<td>0.995</td>
<td>0.978</td>
<td>0.873</td>
</tr>
</tbody>
</table>

Source: Campbell et al., 1997 and the author’s own calculations.
Note: μ_m = 7% and σ_m = 18% in each scenario.

As an example, let us walk through the computations when N = 10 and T = 360 in scenario A. Under the null hypothesis, J_4 has an exact central F distribution with 10 degrees of freedom in the numerator and 349 degrees of freedom in the denominator. In this distribution a size of 5% corresponds to a critical value of 1.8579. Given the Sharpe ratios of the tangency and the market portfolio in scenario A, we have that δ = 14.4805. Inserting these values into (3.61) we get a non-central F distribution under H_A. In this alternative distribution, the probability of exceeding 1.8579 is 20.7% (Campbell et al., 1997).

Generally it can be noted that the power grows from A to B to C. It makes sense that the probability of rejecting the efficiency of the market portfolio increases when its distance to the tangency portfolio increases. Also, note that the power is increasing in T and decreasing in N. Thus, as was the case when evaluating the true size of the asymptotic tests, it is attractive to analyze the behavior of a small number of portfolios over a long time horizon. According to table 3.2, very high powers can be attained when N goes towards 1. In reality, this is unlikely because the Sharpe ratio of the tangency portfolio in the example is probably overstated for smaller values of N. Intuitively, when the number of portfolios is reduced, there are less diversification possibilities. This increases the variance which in turn reduces the power through the noncentrality parameter δ. Thus, there is a tradeoff to consider when choosing N. We cannot determine the optimal value precisely. However, Campbell et al. (1997) suggest that we should probably not include more than 10 portfolios in empirical studies. The following result summarizes section 3.1.4:

Result 3.10: The power of the standard tests is highest when testing data from many periods and a small number of portfolios. The number of portfolios should not be too small though, as the power will start decreasing eventually.
3.5 The assumptions of IID and jointly normal returns

To apply ML, we had to specify a joint pdf of excess returns. We assumed that excess returns are jointly multivariate normal conditional on the market risk premia. That strategy could be unfortunate, as the assumed distribution may be a bad description of the reality (Hall, 2009). If, for example, the joint pdf is incorrect, the foundation upon which we derived the estimators is arguably lost. There is a rich literature indicating that security returns are not normally distributed. Fama (1965) depicts the excess kurtosis consistent with leptokurtic distributions of daily stock returns. Graphically, the empirically observed distribution has fatter tails than you would expect from a normal distribution. Blattberg and Gonedes (1974) demonstrate that the heavier tailed student t distribution is a rather good description of the observed daily stock returns. Affleck-Graves and McDonald (1989) reports significant skewness and excess kurtosis in monthly return data. However, it has become somewhat accepted that monthly returns are “fairly” normally distributed. As the time scale is increased, returns are generally more and more normal (Cont, 2001). All in all, there is some evidence against normality of security returns. Fortunately, normality is only a sufficient, not necessary, condition for the foundation of CAPM: mean-variance analysis. We have only adopted this assumption for statistical testing purposes. Owen and Rabinovitch (1983) document that MVA makes sense if the distribution of returns belong to the class of elliptical distributions. Chamberlain (1983) reaches similar conclusions. An example is the above mentioned student t distribution. In particular, the separation theorem presented in chapter 2 holds if returns follow and elliptical distribution. This theorem was central in the derivation of the CAPM. Summing up, the normality assumption may not be a good description of reality, but it is not mandatory anyway.

Up until now, we have also assumed that excess returns are independently and identically distributed over time. The notion of identically distributed implies homoskedasticity in the time series dimension. Independently means that there is no autocorrelation both linearly and nonlinearly. There is a vast amount of empirical evidence against these assumptions being correct. In liquid markets, however, there is little evidence of significant linear autocorrelation. On the other hand, autocorrelation seems to increase with the time scale as some linear dependence has been found in weekly and monthly data. Furthermore, nonlinear forms of autocorrelation have been documented i.e. in the form of volatility clustering (Cont, 2001). It is commonly known that the volatility of the stock markets changes over time. French, Schwert and Stambaugh (1987) report considerable fluctuation in return volatility. The popularity of models of changing conditional variance is itself an indicator that volatility is not temporally constant (Nelson, 1991). It seems that the notion of IID is not realistic. However, it is a prerequisite for including several periods in empirical tests of a one period model as the CAPM. To see why, recall from chapter 2 that investors maximize their end of period wealth according to MVA. Their portfolio choice is based on the one period means and variances (Wälti, 2007). If returns are non-IID, we cannot consider the different t’s as one, whole stationary period, and the simple analysis of chapter 2 is invalidated. Thus, the CAPM is theoretically wrong by construction if we do not assume IID returns. Nonetheless, as the if the CAPM is a good description of reality, it is still very relevant for practical purposes.

If the returns are not IID and jointly normal, the inference from the statistics developed so far may be unreliable. MacKinlay and Richardson (1991) show that deviations from this assumption can lead to biases in both directions. Combining this point with the above discussion, it should be

---

15 Volatility clustering is the phenomenon that large price changes are followed by more large price changes.
interesting to relax the distributional assumptions. Consequently, we now introduce an alternative estimation method. In contrast to ML, we do not have to specify a joint pdf. Specifically, we do not have to assume joint normality of excess returns. Also, it is robust to autocorrelation and heteroskedasticity (Campbell et al., 1997). The method is called Generalized Method of Moments (GMM) and is based on conditions of the population moments. We only have to assume that excess returns are covariance (weakly) stationary and ergodic. Covariance stationarity implies that the mean and auto covariance are finite and do not depend on the time index \( t \). In an ergodic stationary process, the existing, finite moments are consistently estimated by the sample moments (Steigerwald, 2011). Also, we assume that the fourth moment of excess returns is finite.

Our analysis will be based on the population moment conditions of the market model (3.5) and (3.7). They are

\[
\begin{align*}
E[\varepsilon_t] &= 0, \\
\text{Cov}[Z_{mt}, \varepsilon_t] &= E[Z_{mt}, \varepsilon_t] = 0.
\end{align*}
\]

Both of them are \( N \times 1 \)-vectors, so there are \( 2N \) moment conditions. To apply GMM, we define the function

\[
f_t(\theta) = h_t \otimes \varepsilon_t \tag{3.62}
\]

where

\[
h_t = \begin{bmatrix} 1 \\ Z_{mt} \end{bmatrix}
\]

and

\[
\theta = \begin{bmatrix} \alpha' \\ \beta' \end{bmatrix}.
\]

Note that we have redefined \( \theta \) to denote the vector of parameters to be estimated. Also, note that \( \varepsilon_t = Z_t - \alpha - \beta Z_{mt} \). Furthermore, under the rather weak assumptions above, \( \varepsilon_t \) can be heteroskedastic and exhibit autocorrelation. The symbol \( \otimes \) refers to the Kronecker product. We can thus specify (3.62) as

\[
f_t(\theta) = \begin{bmatrix} 1 \\ Z_{mt} \end{bmatrix} \otimes \varepsilon_t = \begin{bmatrix} (Z_t - \alpha - \beta Z_{mt}) \\ Z_{mt}(Z_t - \alpha - \beta Z_{mt}) \end{bmatrix}, \tag{3.63}
\]

where \( f_t(\theta) \) is a \( 2N \times 1 \)-vector. Using (3.63), we can write the moment conditions above as

\[
E[f_t(\theta_0)] = 0, \tag{3.64}
\]

where \( \theta_0 \) denote the vector of true parameters. These conditions are more generally known as orthogonality conditions as they state that the vector of instruments \( h_t \) is uncorrelated with the vector of errors. The true expected value is of course unknown, so we use the sample counterpart

\[
g_T(\theta) = \frac{1}{T} \sum_{t=1}^{T} f_t(\theta). \tag{3.65}
\]

In the GMM framework, we use (3.65) to set up the function

\[
Q_T(\theta) = g_T(\theta)'Wg_T(\theta), \tag{3.66}
\]

where \( W \) is a \( 2N \times 2N \)-matrix that weights the orthogonality conditions. GMM seeks to find the parameter estimates that, given the observations, satisfy the specified orthogonality conditions as closely as possible. This is equivalent to finding the estimates that minimize \( Q_T(\theta) \). However, if the number of orthogonality conditions equal the number of parameters to be estimated, the system
is said to be *exactly identified*. Then, a set of unique estimators can be determined so that the average of the sample moments (3.65) exactly equals zero. This is the case here, as there are two parameters \( N \) times in the vector \( \theta \) and two population moment conditions \( N \) times cross-sectionally in the market model.

Finding the optimal estimates is a standard quadratic minimization problem. We therefore differentiate (3.66) with respect to the parameter vector, set the result equal to zero, and solve for the estimators. We have

\[
\frac{\partial Q_T(\theta)}{\partial \theta}(\hat{\theta}) = D_T(\hat{\theta})' W g_T(\hat{\theta}) = 0, \tag{3.67}
\]

where

\[
D_T(\theta) = \frac{\partial g_T(\theta)}{\partial \theta'}.
\]

The term \( D_T(\theta) \) is a \( 2N \times 2N \)-matrix, and the derivative in (3.67) is a \( 2N \times 1 \)-vector. In (3.67), we consequently have a set of \( 2N \) equations to solve with \( 2N \) unknown parameters contained in the vectors \( \hat{\alpha} \) and \( \hat{\beta} \). As the system is exactly identified, the solution becomes independent of the weights. To see why, note that the moment conditions can be satisfied exactly using only (3.65), so there is no need to weight them against each other. Thus, the minimum of zero for \( Q_T(\hat{\theta}) \) is obtained regardless of the \( W \) used. As a consequence, the parameter estimators turn out to be equivalent to the ML/OLS estimators derived before. They are

\[
\hat{\alpha} = \bar{\mu} - \hat{\beta} \mu_m \tag{3.68}
\]

and

\[
\hat{\beta} = \frac{\sum_{t=1}^{T}(Z_t - \bar{\mu})(Z_{mt} - \bar{\mu}_m)}{\sum_{t=1}^{T}(Z_{mt} - \bar{\mu}_m)^2}. \tag{3.69}
\]

A central contribution of this GMM procedure lies in the variance-covariance matrix of the estimators. It is robust to returns not being IID. Thus they can be applied to form a test that is not affected by heteroskedasticity and autocorrelation. Also, note that we have assumed nothing about normality. Then, the test statistic we build can serve as a robustness check of the inference from \( f_0 - f_3 \). The variance-covariance matrix of \( \hat{\theta} \) is

\[
V = [D_0'S_0^{-1}D_0]^{-1}, \tag{3.70}
\]

where

\[
D_0 = E \left[ \frac{\partial g_T(\theta)}{\partial \theta'} \right]
\]

and

\[
S_0 = \sum_{t=\infty}^{+\infty} E \left[ \frac{1}{T} f_t(\theta)f_{t-1}(\theta)' \right].
\]

The estimator \( \hat{\theta} \) has a normal distribution asymptotically. As before, the ML/OLS components are unbiased. We can therefore write the distribution of the GMM estimator as
\[
\hat{\theta} \sim N(\theta, \tilde{V}/T) = N\left(\theta, \frac{1}{T} [\tilde{D}_0 \tilde{S}_0^{-1} \tilde{D}_0]^{-1}\right).
\] (3.71)

We now only need to find consistent estimators \( \tilde{D}_0 \) and \( \tilde{S}_0 \) to construct the test statistic. Using (3.63) and (3.65), we have

\[
\tilde{D}_0 = E\left[\frac{\partial g_r(\theta)}{\partial \theta'}\right] = \\
= \begin{bmatrix}
\frac{1}{T} \sum_{t=1}^{T} (z_t - \alpha - \beta z_{mt}) & \frac{1}{T} \sum_{t=1}^{T} (z_t - \alpha - \beta z_{mt}) \\
\frac{1}{T} \sum_{t=1}^{T} z_{mt} (z_t - \alpha - \beta z_{mt}) & \frac{1}{T} \sum_{t=1}^{T} z_{mt} (z_t - \alpha - \beta z_{mt})
\end{bmatrix}
\]
\[
= -\left[\frac{1}{\mu_m} \begin{bmatrix}
\mu_m \\
\nu_m \\
\end{bmatrix} \right] \otimes I_N,
\] (3.72)

where \( I_N \) is a \( N \times N \)-identity matrix. The third equality is straightforward. Just recall from (3.50) that we can write \( (\sigma_m^2 + \mu_m^2) = \frac{1}{T} \sum_{t=1}^{T} z_{mt}^2 \). The estimator \( \tilde{D}_0 \) can be created by substituting for the sample counterparts in (3.72). We can thus reuse the ML estimators (3.29) and (3.34) to get

\[
\tilde{D}_0 = -\left[\frac{1}{\mu_m} \begin{bmatrix}
\mu_m \\
\nu_m \\
\end{bmatrix} \right] \otimes I_N.
\] (3.73)

Considering the term \( S_0 \) we need a way to work around the infinite number of terms, as one can only estimate a finite number of terms in a finite sample. Furthermore, we need to be certain that the estimator of \( S_0 \) is positive definite, as the variance \( V \) must be positive definite. Newey and West developed such an estimator in 1987. It can be defined as

\[
\tilde{S}_0(q, \theta) = \bar{r}_{0,T}(\theta) + \sum_{j=1}^{q} \left(\frac{q-j}{q}\right) \left(\bar{r}_{j,T}(\theta) + \bar{r}'_{j,T}(\theta)\right),
\] (3.74)

where

\[
\bar{r}_{j,T}(\theta) = \frac{1}{T} \sum_{t=j+1}^{T} f_t(\theta) f_{t-j}'(\theta).
\] (3.75)

Note that the autocorrelation of \( f_t(\theta) \) is assumed to disappear in the \( q \)’th lag. Here Newey and West (1994) suggest the following estimate \( q = 4(T/100)^{2/9} \) rounded down.

In (3.71), we only got a variance-covariance matrix of \( \hat{\theta} \). In order to specify a test statistic for testing the null hypothesis, we need \( \text{Var}(\hat{\alpha}) \). We therefore define \( \hat{\alpha} \) in terms of \( \hat{\theta} \) by writing \( \hat{\alpha} = R \hat{\theta} \), where \( R = (1 \ 0) \otimes I_N \). Then, we get a consistent, robust estimator of the variance of \( \hat{\alpha} \). We have
\[ \text{Var}(\hat{\alpha}) = \text{Var}(\mathbf{R}\hat{\theta}) = \mathbf{R}\text{Var}(\hat{\theta})\mathbf{R}' = \mathbf{R} \frac{1}{T} \left[ \mathbf{D}_0\mathbf{S}_0^{-1}\mathbf{D}_0 \right]^{-1} \mathbf{R}'. \] (3.76)

Note that the asymptotic equality follows from the asymptotic distribution in (3.71). Now we have the necessary tools to build a Wald type test similar to (3.40). It has a chi-square distribution with \( N \) degrees of freedom asymptotically. This can be summarized into the following result (Campbell et al., 1997):

**Result 3.11:** The robust Wald test of the null hypothesis (3.8) can be defined as

\[ J_T = T\hat{\alpha}' \left[ \mathbf{R} \left[ \hat{\mathbf{D}}_0\hat{\mathbf{S}}_0^{-1}\hat{\mathbf{D}}_0 \right]^{-1} \mathbf{R}' \right]^{-1} \hat{\alpha} \sim \chi^2_N. \] (3.77)

### 3.6 Cross-Sectional Regressions

Up until now we have developed statistics for testing the mean-variance efficiency of the market portfolio. Now, we will briefly consider tests of the two other implications of the CAPM: beta can explain the variation in the cross-sections of expected excess returns perfectly, and the risk premium in the market is positive. To do this we apply what is known as the Fama-MacBeth procedure originally developed in 1973. We will continue the assumption from section 3.1.2 that excess returns are normally distributed and temporally IID. Also, in this procedure, we assume that the true value of the betas are known. This assumption will be discussed later. Conversely, we do not include observations of market risk premium, as we wish to test whether the population value can be assumed positive. In the Fama-MacBeth procedure, we estimate \( T \) cross-sectional regression models. For the model in the \( t \)'th cross section we have

\[ \mathbf{Z}_t = \gamma_0 \mathbf{t} + \gamma_1 \mathbf{t}_m + \mathbf{\eta}_t, \] (3.78)

where \( \mathbf{\eta}_t \) is a \( N \times 1 \)-vector of error terms, \( \mathbf{t} \) is a \( N \times 1 \)-vector of ones, and the rest is defined as previously. The procedure has two steps. First, having \( T \) vectors of observed betas and excess returns, we estimate (3.78) cross-section by cross-section using OLS. The result is \( T \) estimates of \( \gamma_0 \) and \( \gamma_1 \). Then, using these estimates, we conduct aggregated tests of the above implications. Let \( E[\gamma_0] = \gamma_0 \) and \( E[\gamma_1] = \gamma_1 \). Then, the above implications can be transformed into the hypotheses:

Against

\[ \text{H}_0: \gamma_0 = 0 \] (3.79)

And

\[ \text{H}_A: \gamma_0 \neq 0, \] (3.80)

Against

\[ \text{H}_0: \gamma_1 = 0 \] (3.81)

And

\[ \text{H}_A: \gamma_1 \neq 0. \] (3.82)

If the CAPM holds, we must fail to reject (3.79) and be able to reject (3.81). Also \( \gamma_1 \) should have a positive estimate. The notion that \( \gamma_0 = 0 \) has a similar interpretation to \( \alpha = 0 \). It says that excess returns are, on average, perfectly related to the betas - nothing else matters. In the second hypothesis, CAPM indicates that \( \gamma_1 > 0 \). As \( \gamma_1 \) is the market risk premium, this is equivalent to saying that investors are rewarded for bearing extra risk. To test the hypotheses we need estimators of \( \gamma_0 \) and \( \gamma_1 \). We use their sample counterparts which are the simple averages of the cross-sectional estimates.
\[
\hat{y}_0 = \frac{1}{T} \sum_{t=1}^{T} \hat{y}_{0t}
\]  
(3.83)

and

\[
\hat{y}_1 = \frac{1}{T} \sum_{t=1}^{T} \hat{y}_{1t}.
\]  
(3.84)

The distribution of the estimators follow from the assumptions regarding the behavior of excess returns. They also have a normal distribution and are IID over time. Furthermore, the variance of the estimators can be written as

\[
\hat{\sigma}_{y_j}^2 = \frac{1}{T(T-1)} \sum_{t=1}^{T} (\hat{y}_{jt} - \hat{y}_j)^2.
\]  
(3.85)

We now have the necessary information to form a test statistic. Note that the above hypotheses are not joint hypotheses since the \(y_j\)'s are scalars. Therefore, we use simple \(t\)-tests. Noting that they will have a student \(t\) distribution with \((T - 1)\) degrees of freedom under the null hypotheses, we have the following results:

**Result 3.12:** The \(t\)-test of the null hypothesis (3.79) can be defined as

\[
\omega(\hat{y}_0) = \frac{\hat{y}_0}{\hat{\sigma}_{y_0}} \sim t(T - 1).
\]  
(3.86)

**Result 3.13:** The \(t\)-test of the null hypothesis (3.81) can be defined as

\[
\omega(\hat{y}_1) = \frac{\hat{y}_1}{\hat{\sigma}_{y_1}} \sim t(T - 1).
\]  
(3.87)

The force of the Fama-MacBeth procedure is that the model allows for more in-depth tests. That is, (3.78) can easily be modified to include other explanatory variables. Then, one can test whether these risk measures have any influence on excess returns, once beta has been controlled for. Thus, we actually test whether beta perfectly describes the variation in the cross-sections of excess returns. To conduct such tests, define the cross-sectional model as

\[
Z_t = \gamma_0 t + \gamma_1 t \beta_m + \gamma_2 t \xi_t + \eta_t,
\]  
(3.88)

where \(\xi_t\) is a \(N \times 1\)-vector of some additional risk measure. Examples include firm characteristics as market capitalization, price-earning-ratios etc. We can thus test the hypothesis \(H_0: \gamma_2 = 0\) by inserting the estimates from (3.88) in the above formulae. If we fail to reject, the additional risk measure has some explanatory power that beta does not capture. This would clearly be evidence against the CAPM.

One obvious drawback to the Fama-MacBeth approach is that we do not know the true market betas. In empirical implementations we use beta estimates from the sample data. This measurement error results in an errors-in-variables problem. One solution to the problem is to increase the precision of the beta estimates by grouping the securities into portfolios. Another solution is the adjustment factor presented by Shanken (1992). The variance of the parameter estimators (3.85) should simply be multiplied by the term \((1 + (\hat{\mu}_m - \hat{y}_0)^2 / \hat{\sigma}_m^2)\) (Campbell et al., 1997).
Chapter 4 Empirical study

"...the proper test of a theory is not the realism of its assumptions but the acceptability of its implications..." (Sharpe, 1964, p. 434).

First, a brief account of the significant historical contributions regarding the empirical validity of the CAPM is given. Then, issues in implementing the tests we have derived are discussed. This will serve as a basis for the data selection decisions which are the topic of the next section. The chapter is concluded by an empirical test of the CAPM.

4.1 Literature review

Since the original papers of Lintner and Sharpe in the mid 1960s, the CAPM has been subject to a huge number of empirical studies. Both time-series and cross-sectional tests have been conducted from the beginning. Among the early tests, Black et al. (1972) and Fama and MacBeth (1973) are perhaps the most notable examples of such tests. Other contributions include Douglas (1968), Friend and Blume (1970), Miller and Scholes (1972), and Blume and Friend (1973), and Stambaugh (1982). The early tests generally reject the Sharpe-Lintner model. They found a positive, linear relation between beta and expected return, but it was not steep enough compared to the SML. The tests found that the intercept was above the risk free rate and that the observed the slope was smaller than the market risk premium. The implication was that low beta portfolios have positive alphas, and that high beta portfolios have negative alphas. This is not consistent with the traditional Sharpe-Lintner version. However, the version without risk-free borrowing and lending developed by Black (1972) could largely explain these deviations. The coefficient on beta was positive and beta seemed to completely explain expected returns in the Black model. The early empirical success of the Black version combined with the simplicity of the model heavily contributed to the popularity of the CAPM.

At the end of the 1970s more critical studies started appearing. Even the Black version could not handle these blows. Collectively known as the anomalies literature, the studies concentrate on the variation in expected return unrelated to beta. Specifically, they find a relationship between firm characteristics and expected return unaccounted for by the CAPM. Basu’s 1977 paper documents a price-earnings effect that invalidates the efficiency of the market portfolio. It shows that low price-earnings ratios are coupled with too high returns compared to the SML and vice versa for high price-earnings ratios. Also, numerous tests have been concerned with the size-effect first reported by Banz (1981). It states that small firms (as measured by market capitalization) have higher returns than the CAPM predicts. Likewise, a leverage effect has been found by Bhandari (1988) while Statman (1980) and Rosenberg et al. (1985) provide evidence that high book-to-market equity ratios are coupled with returns above what beta predicts. These effects are confirmed in a more recent study by Fama and French (1992). In this paper they also argue that the relation between expected return and beta may be even flatter than the one reported by the early empirical literature. This point was however challenged in 1995 by Kothari et al. (Fama and French, 2004). One could argue that the reported effects seem to lack a theoretical anchor. Consequently, one cannot rule out the possibility that the anomalies literature is subject to data-snooping and selection biases. Data-snooping biases can occur when researchers are guided in their study by previous results using the same data. For example, when hypotheses etc. are formed after looking at the data. The vast amount
of empirical tests of the CAPM contributes to the risk of data-snooping biases. Sample selection biases can arise when a certain part of the population is systematically avoided when collecting data. This may occur due to data availability issues. An example is the survivorship bias reported by Kothari et al. (1995) that can arise when considering book-to-market ratios (Campbell et al., 1997).

4.2 Issues in implementing the tests
Before drawing the sample, we seek to answer the questions that arise when implementing the tests of the CAPM. The relevant topics include the choice of proxies for the market portfolio and the risk free rate, length of the period and sampling frequency, and the choice of portfolios.

4.2.1 The choice of proxies
A central element in the CAPM is the market portfolio. We have argued that testing whether the intercepts in the time-series market model are zero, is equivalent to testing the mean-variance efficiency of the market portfolio. The market portfolio consists of all risky assets including human capital, real estate, bonds etc. In his important paper, Roll (1977) underlined that what we actually test is the proxy for the market portfolio. Most proxies typically only include stocks, moreover only American stocks. Consequently, empirical studies really only reject the efficiency of the proxy they use, not the CAPM. However, there is some evidence that this is merely a theoretical argument. It has been shown that inferences are not dependent on whether the proxy only includes stocks, or also contains bonds and real-estate. Also, it has been documented that as long as there is a high correlation (above 0.70) between the proxy and the market portfolio, we can trust a rejection of the CAPM (Campbell et al., 1997).

It follows from result 2.2 e that the market portfolio is by definition a value-weighted portfolio, where every asset is weighted by its share of the total market value. Thus, it is theoretically consistent to use a value-weighted portfolio for the market proxy. On the other hand, Bartholdy and Peare (2005), showed that an equal weighted index results in a better coefficient of determination R² in a Fama-Macbeth type cross-sectional regression than value-weighted indexes do. Thus, a equal-weighted index may perform better in real world applications of the CAPM. Both choices are represented in the literature.

Theoretically, one should include dividends when testing CAPM. The CAPM computes the expected reward for bearing systematic risk. This reward must logically be the total return, not just capital appreciation. Lintner (1965 b) also explicitly mentions dividends in his development of the model. However, concerning the choice of market portfolio proxy, Bartholdy and Pearson (2005) show that indexes with and without dividends are very highly correlated. Thus, if a dividend-adjusted index is not available, it probably does not affect the outcome.

Probably the most widely used proxy for the market portfolio is the Standard & Poors 500 composite (S&P 500) index (Campbell et al., 1997). However, the Center for Research in Security Prices (CRSP) market index is also widely used in more recent contributions. Notable examples include Gibbons et al. (1989), MacKinlay and Richardson (1991), and Fama and French (2004).

The proxy for the risk free rate is typically the 1-month US Treasury Bill (Fama and French, 2004). This proxy virtually has no default risk, as it is backed by the US government. Furthermore, due to its short maturity, it is also practically free of interest rate risk.
4.2.2 Period length and sampling frequency
The CAPM model relates expected (ex-ante) returns to beta. Of course, we cannot observe expected returns. However it is possible to test the CAPM using realized (ex-post) mean returns (Black et al., 1972).

Shortly put, the choice of period length is a balance between non-constancy of beta and statistical power (Fama and MacBeth, 1973). According to result 3.9 and 3.10 the true size and power of the tests are optimized when including many periods in the sample. On the other hand, the true betas are likely to change over time. This introduces a bias in the long period estimates. To solve the problem, one can use a high sampling frequency to get a lot of observations over time but still keep the period relatively short. On the other hand, there is probably too much noise in daily data which reduces the efficiency of the estimates. The standard choice is 5 years of monthly data (Bartholdy and Peare, 2005; Groenewold and Fraser, 2001).

4.2.3 Portfolio construction
In some early tests of the CAPM individual stocks were used. Later, it has been common to use portfolios as they increase the precision of the beta estimates. This reduces the errors-in-variables problem discussed above. Furthermore, they are straightforward to use, as the expected return and beta of a portfolio are just the weighted averages of the individual returns and betas respectively. However, this raises the problem of how to construct the portfolios. Also, grouping assets into portfolios reduces the range of the betas. A common solution is to sort assets using pre-estimated betas and then group them into portfolios based on their ranks. This ensures a wider spread in the observations (Fama and French, 2004). However, one can also sort on size as this ensures a wide spread in average returns and betas as well (Fama and French, 1992).

4.3 Data selection
For the empirical tests, we use data from the New York Stock Exchange, the American Stock Exchange (Amex), and NASDAQ. First, data for many decades are easily accessible. Second, they feature some of the world’s largest and most liquid markets (WFE, 2009). Following the recommendations from the sections on true size (3.1.3), power (3.1.4), and portfolio construction (4.2.3), the eligible stocks of the three exchanges are assigned to 10 value-weighted portfolios based on their size. Specifically, at the end of each quarter, eligible companies on the NYSE are sorted on market capitalization and then divided into 10 deciles of equal populations. The companies with the largest capitalization in each NYSE decile serve as the breakpoints when assigning all the sampled companies. Portfolio 1 contains the largest companies, portfolio 2 the next largest, and so forth. Note that NASDAQ was first included from April 1982. The returns recorded are total returns, so dividends are included as demanded by the CAPM. As indicated above, the returns are computed from a value-weighted portfolio of the securities in each decile.

Consistent with section 4.2.1, we also choose the theoretical consistent value-weighted proxy of the market portfolio. Specifically, we use the CRSP value weighted basket of American stocks with dividends. It includes the issues on the NYSE, Amex, and NASDAQ. To check the robustness of the inference to the market portfolio proxy used, we also employ the equivalent CRSP equal weighted portfolio. For the proxy of the risk free rate, we use the 30-day US Treasury Bill as

---

16 Unit Investment Trusts, Closed-End funds, Real Estate Investment Trusts, Americus Trusts, Foreign Stocks, and American Depository Receipts are all excluded.
recommended above. All data is collected from the CRSP tapes of Wharton Research Data Services (WRDS, 2011).

We use 30 years of monthly data from January 1981 to December 2010. This way we can also test the model’s performance on newer data, as the main part of the empirical literature deals with samples from before the 1990s. Consistent with the previous section we divide the period into 6 subperiods of 5 years. To get more flexibility regarding the tradeoff between statistical power and non-stationarity of the true parameter values, we also evaluate subperiods of 10 years. As a result we have a sample with the dimensions $N = 10$ and $T = 360$ for the whole period, $T = 120$ for the 10 year periods, and $T = 60$ for each 5 year subperiod.

4.4 Empirical tests

First, a preliminary evaluation of $\hat{\alpha}$ and $\hat{\beta}$ for the 10 portfolios is provided. Then, the results from the tests of $H_0$: $\alpha = 0$ are presented followed by the results from the tests of $H_0$: $\gamma_0 = 0$ and $H_0$: $\gamma_1 = 0$. Lastly, we conduct a graphical interpretation of the tests.

4.4.1 Parameter estimates of the market model

The parameter estimates from the market model (3.1.4) are reported in table 4.1 for the entire period and the 10 year subperiods. Note that the beta estimates are highest in the last 10 years for 7 out of 10 portfolios. Thus, it seems that the responsiveness of the excess returns to market conditions is more pronounced in the last decade of the sample period.

Furthermore, it seems that the baskets of small companies are somewhat more sensitive to swings in economic activity than the portfolios of large capitalizations. The beta estimates are increasing from portfolio 1 through 8. Only the two smallest deciles break the trend. One feasible explanation is that the largest companies comprise a considerable part of the proxy for the market portfolio. It should also be noted that the portfolios with betas above 1 appear to have a positive Jensen’s alpha in the overall 30 year period. This is not consistent with the previous evidence that the empirical relationship between return and beta is flatter than the theoretical relationship. More on that later. In table 4.2 the same information is provided for the 5 year subperiods. For a fixed portfolio, major fluctuations in the parameter estimates are evident. Especially for the portfolios of smaller firms, it seems unreasonable to assume that the true alphas and betas are unchanged through the sample period.
Table 4.1: Parameter estimates from the market model of the 10 size-sorted portfolios in the 30 year period and 10 year subperiods.

<table>
<thead>
<tr>
<th>Portfolio</th>
<th>Parameter*</th>
<th>30 year period</th>
<th>10 year subperiods</th>
<th>10 year subperiods</th>
<th>10 year subperiods</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$\hat{a}$</td>
<td>-0.0006</td>
<td>0.0113</td>
<td>0.0090</td>
<td>-0.0272</td>
</tr>
<tr>
<td></td>
<td>$\beta$</td>
<td>0.9394</td>
<td>0.9464</td>
<td>0.9874</td>
<td>0.8971</td>
</tr>
<tr>
<td>2</td>
<td>$\hat{a}$</td>
<td>0.0111</td>
<td>0.0074</td>
<td>0.0064</td>
<td>0.0228</td>
</tr>
<tr>
<td></td>
<td>$\beta$</td>
<td>1.0178</td>
<td>1.0425</td>
<td>0.9798</td>
<td>1.0217</td>
</tr>
<tr>
<td>3</td>
<td>$\hat{a}$</td>
<td>0.0113</td>
<td>0.0077</td>
<td>-0.0027</td>
<td>0.0323</td>
</tr>
<tr>
<td></td>
<td>$\beta$</td>
<td>1.0805</td>
<td>1.0761</td>
<td>1.0465</td>
<td>1.1115</td>
</tr>
<tr>
<td>4</td>
<td>$\hat{a}$</td>
<td>0.0121</td>
<td>0.0070</td>
<td>-0.0100</td>
<td>0.0412</td>
</tr>
<tr>
<td></td>
<td>$\beta$</td>
<td>1.0886</td>
<td>1.0854</td>
<td>1.0689</td>
<td>1.1112</td>
</tr>
<tr>
<td>5</td>
<td>$\hat{a}$</td>
<td>0.0124</td>
<td>-0.0005</td>
<td>-0.0160</td>
<td>0.0514</td>
</tr>
<tr>
<td></td>
<td>$\beta$</td>
<td>1.0934</td>
<td>1.0510</td>
<td>1.1248</td>
<td>1.1210</td>
</tr>
<tr>
<td>6</td>
<td>$\hat{a}$</td>
<td>0.0115</td>
<td>-0.0084</td>
<td>-0.0026</td>
<td>0.0483</td>
</tr>
<tr>
<td></td>
<td>$\beta$</td>
<td>1.1542</td>
<td>1.1005</td>
<td>1.1386</td>
<td>1.2206</td>
</tr>
<tr>
<td>7</td>
<td>$\hat{a}$</td>
<td>0.0099</td>
<td>-0.0193</td>
<td>-0.0062</td>
<td>0.0623</td>
</tr>
<tr>
<td></td>
<td>$\beta$</td>
<td>1.1598</td>
<td>1.1369</td>
<td>1.0958</td>
<td>1.2309</td>
</tr>
<tr>
<td>8</td>
<td>$\hat{a}$</td>
<td>0.0145</td>
<td>-0.0197</td>
<td>-0.0134</td>
<td>0.0829</td>
</tr>
<tr>
<td></td>
<td>$\beta$</td>
<td>1.1863</td>
<td>1.1331</td>
<td>1.1356</td>
<td>1.2806</td>
</tr>
<tr>
<td>9</td>
<td>$\hat{a}$</td>
<td>0.0104</td>
<td>-0.0444</td>
<td>0.0028</td>
<td>0.0831</td>
</tr>
<tr>
<td></td>
<td>$\beta$</td>
<td>1.1556</td>
<td>1.0752</td>
<td>1.0734</td>
<td>1.2927</td>
</tr>
<tr>
<td>10</td>
<td>$\hat{a}$</td>
<td>0.0152</td>
<td>-0.0740</td>
<td>0.0192</td>
<td>0.1239</td>
</tr>
<tr>
<td></td>
<td>$\beta$</td>
<td>1.0620</td>
<td>1.0065</td>
<td>0.8494</td>
<td>1.2623</td>
</tr>
</tbody>
</table>

Note: * The $\hat{a}$’s are annualized.

This is consistent with the guidelines given earlier and suggests that we should interpret the test results from the entire period only with care. On the other hand, it seems more likely that the estimates are constant in each 10 year subperiod even though there are examples of substantial intra-period changes here also. However, as mentioned before, we also lose statistical power when moving from the 30 year period to the 5 subperiods. Small samples are more prone to sampling error, and the efficiency of the estimates is reduced compared to larger samples. Gonedes (1973) provides a useful discussion. Reviewing the arguments, we should probably consider all three alternatives when evaluating the empirical performance of the CAPM.
4.4.2 Time-series tests of the intercept

Recall that the joint hypothesis from chapter 3 was \( H_0: \alpha = 0 \) against \( H_A: \alpha \neq 0 \). In table 4.4 the results of the tests related to this hypothesis are presented. First, notice that the true-size problems of the asymptotic tests are exposed. The Wald-test \( J_0 \) and the Likelihood-ratio test \( J_2 \) reject \( H_0 \) at the common 5% significance level on several occasions where the exact \( F \)-test \( J_1 \) does not. Conversely, the great performance of the finite-sample correction of the likelihood-ratio test is underlined as the p-values of \( J_1 \) and \( J_3 \) are almost identical in each period. Consequently, we will concentrate on drawing inference from \( J_1 \) and the robust test \( J_7 \) in the following analysis.

**Table 4.2: Parameter estimates from the market model of the 10 size-sorted portfolios in the 5 year subperiods.**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>( \hat{\alpha} )</td>
<td>-0.0019</td>
<td>0.0248</td>
<td>-0.0067</td>
<td>0.0262</td>
<td>-0.0408</td>
<td>-0.0134</td>
</tr>
<tr>
<td></td>
<td>( \beta )</td>
<td>0.9225</td>
<td>0.9609</td>
<td>0.9689</td>
<td>0.9939</td>
<td>0.9313</td>
<td>0.8736</td>
</tr>
<tr>
<td>2</td>
<td>( \hat{\alpha} )</td>
<td>0.0131</td>
<td>0.0018</td>
<td>0.0061</td>
<td>-0.0007</td>
<td>0.0415</td>
<td>0.0038</td>
</tr>
<tr>
<td></td>
<td>( \beta )</td>
<td>1.0119</td>
<td>1.0618</td>
<td>1.0723</td>
<td>0.9480</td>
<td>0.9764</td>
<td>1.0529</td>
</tr>
<tr>
<td>3</td>
<td>( \hat{\alpha} )</td>
<td>0.0167</td>
<td>-0.0013</td>
<td>0.0099</td>
<td>-0.0196</td>
<td>0.0497</td>
<td>0.0151</td>
</tr>
<tr>
<td></td>
<td>( \beta )</td>
<td>1.0869</td>
<td>1.0695</td>
<td>1.1001</td>
<td>1.0281</td>
<td>1.1559</td>
<td>1.0823</td>
</tr>
<tr>
<td>4</td>
<td>( \hat{\alpha} )</td>
<td>0.0174</td>
<td>-0.0037</td>
<td>0.0060</td>
<td>-0.0283</td>
<td>0.0659</td>
<td>0.0167</td>
</tr>
<tr>
<td></td>
<td>( \beta )</td>
<td>1.1149</td>
<td>1.0674</td>
<td>1.0976</td>
<td>1.0589</td>
<td>1.1366</td>
<td>1.0951</td>
</tr>
<tr>
<td>5</td>
<td>( \hat{\alpha} )</td>
<td>0.0289</td>
<td>-0.0297</td>
<td>0.0410</td>
<td>-0.0735</td>
<td>0.0548</td>
<td>0.0481</td>
</tr>
<tr>
<td></td>
<td>( \beta )</td>
<td>1.0199</td>
<td>1.0715</td>
<td>1.1313</td>
<td>1.1222</td>
<td>1.1300</td>
<td>1.1150</td>
</tr>
<tr>
<td>6</td>
<td>( \hat{\alpha} )</td>
<td>0.0391</td>
<td>-0.0559</td>
<td>0.0307</td>
<td>-0.0359</td>
<td>0.0750</td>
<td>0.0214</td>
</tr>
<tr>
<td></td>
<td>( \beta )</td>
<td>1.0867</td>
<td>1.1106</td>
<td>1.1396</td>
<td>1.1380</td>
<td>1.1988</td>
<td>1.2363</td>
</tr>
<tr>
<td>7</td>
<td>( \hat{\alpha} )</td>
<td>0.0143</td>
<td>-0.0532</td>
<td>0.0280</td>
<td>-0.0425</td>
<td>0.0915</td>
<td>0.0334</td>
</tr>
<tr>
<td></td>
<td>( \beta )</td>
<td>1.1680</td>
<td>1.1185</td>
<td>1.1232</td>
<td>1.0862</td>
<td>1.2651</td>
<td>1.2089</td>
</tr>
<tr>
<td>8</td>
<td>( \hat{\alpha} )</td>
<td>0.0307</td>
<td>-0.0709</td>
<td>0.0084</td>
<td>-0.0379</td>
<td>0.1259</td>
<td>0.0403</td>
</tr>
<tr>
<td></td>
<td>( \beta )</td>
<td>1.2133</td>
<td>1.0845</td>
<td>1.1692</td>
<td>1.1239</td>
<td>1.3476</td>
<td>1.2371</td>
</tr>
<tr>
<td>9</td>
<td>( \hat{\alpha} )</td>
<td>0.0214</td>
<td>-0.1107</td>
<td>0.0546</td>
<td>-0.0467</td>
<td>0.1379</td>
<td>0.0284</td>
</tr>
<tr>
<td></td>
<td>( \beta )</td>
<td>1.1135</td>
<td>1.0534</td>
<td>1.0471</td>
<td>1.0821</td>
<td>1.3200</td>
<td>1.2764</td>
</tr>
<tr>
<td>10</td>
<td>( \hat{\alpha} )</td>
<td>0.0057</td>
<td>-0.1544</td>
<td>0.0977</td>
<td>-0.0600</td>
<td>0.2162</td>
<td>0.0313</td>
</tr>
<tr>
<td></td>
<td>( \beta )</td>
<td>1.0780</td>
<td>0.9642</td>
<td>0.8606</td>
<td>0.8451</td>
<td>1.2298</td>
<td>1.2877</td>
</tr>
</tbody>
</table>

Note: * The \( \hat{\alpha} \)'s are annualized.
Table 4.4: Tests statistic values and p-values for the tests of the null-hypothesis $\alpha = 0$ using the CRSP value-weight proxy of the market portfolio.

<table>
<thead>
<tr>
<th>Time</th>
<th>$J_0$</th>
<th>p-value</th>
<th>$J_1$</th>
<th>p-value</th>
<th>$J_2$</th>
<th>p-value</th>
<th>$J_3$</th>
<th>p-value</th>
<th>$J_7$</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>30 year period</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>01/1981-12/2010</td>
<td>12.287</td>
<td>0.266</td>
<td>1.191</td>
<td>0.295</td>
<td>12.082</td>
<td>0.280</td>
<td>11.847</td>
<td>0.295</td>
<td>11.286</td>
<td>0.336</td>
</tr>
<tr>
<td>10 year subperiods</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>01/1981-12/1990</td>
<td>20.075</td>
<td>0.029</td>
<td>1.823</td>
<td>0.065</td>
<td>18.562</td>
<td>0.046</td>
<td>17.479</td>
<td>0.064</td>
<td>23.138</td>
<td>0.010</td>
</tr>
<tr>
<td>01/1991-12/2000</td>
<td>14.037</td>
<td>0.171</td>
<td>1.275</td>
<td>0.253</td>
<td>13.275</td>
<td>0.209</td>
<td>12.500</td>
<td>0.253</td>
<td>19.566</td>
<td>0.034</td>
</tr>
<tr>
<td>01/2001-12/2010</td>
<td>16.084</td>
<td>0.097</td>
<td>1.461</td>
<td>0.164</td>
<td>15.094</td>
<td>0.129</td>
<td>14.213</td>
<td>0.163</td>
<td>17.559</td>
<td>0.063</td>
</tr>
<tr>
<td>5 year subperiods</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>01/1981-12/1985</td>
<td>31.139</td>
<td>0.001</td>
<td>2.543</td>
<td>0.015</td>
<td>25.082</td>
<td>0.005</td>
<td>22.156</td>
<td>0.014</td>
<td>50.212</td>
<td>0.000</td>
</tr>
<tr>
<td>01/1986-12/1990</td>
<td>23.244</td>
<td>0.010</td>
<td>1.898</td>
<td>0.068</td>
<td>19.646</td>
<td>0.033</td>
<td>17.354</td>
<td>0.067</td>
<td>20.619</td>
<td>0.024</td>
</tr>
<tr>
<td>01/1991-12/1995</td>
<td>15.810</td>
<td>0.105</td>
<td>1.291</td>
<td>0.262</td>
<td>14.033</td>
<td>0.171</td>
<td>12.396</td>
<td>0.259</td>
<td>14.570</td>
<td>0.149</td>
</tr>
<tr>
<td>01/1996-12/2000</td>
<td>19.442</td>
<td>0.035</td>
<td>1.588</td>
<td>0.138</td>
<td>16.841</td>
<td>0.078</td>
<td>14.876</td>
<td>0.137</td>
<td>28.445</td>
<td>0.002</td>
</tr>
<tr>
<td>01/2001-12/2005</td>
<td>26.902</td>
<td>0.003</td>
<td>2.197</td>
<td>0.034</td>
<td>22.226</td>
<td>0.014</td>
<td>19.633</td>
<td>0.033</td>
<td>30.360</td>
<td>0.001</td>
</tr>
<tr>
<td>01/2006-12/2010</td>
<td>10.347</td>
<td>0.411</td>
<td>0.845</td>
<td>0.589</td>
<td>9.545</td>
<td>0.481</td>
<td>8.432</td>
<td>0.587</td>
<td>14.208</td>
<td>0.164</td>
</tr>
</tbody>
</table>

Note: For $J_7$ we use $q = 5$ for the 30 year period, $q = 4$ for the 10 year subperiods, and $q = 3$ for the 5 year subperiods.

For the entire 30 year period, we fail to reject the null hypothesis at the common levels of significance. The p-value of the exact $F$-test is 0.295. The alphas reported above of the 30 year period are not jointly significantly different from zero. Consequently, we cannot reject the mean-variance efficiency of the CRSP value-weighted proxy. In other words, we fail to reject the exact linear relationship expressed in the model. Consistent with the CAPM, this indicates that beta completely explains expected excess returns and that no other parameter is needed. This disputes the evidence of a size-effect, as the portfolios are size-sorted. Recall, that alpha is the difference between realized mean return and the expected return predicted by CAPM. If there was a size-effect, one would expect abnormal returns (in both directions) on the portfolios, however, in the last 30 years, we have so far found no evidence of that. This inference is confirmed in the 10 year subperiods. However, the $p$-values suggest that it may be appropriate to differ between the subperiods. Looking at the subperiods, we find evidence against the CAPM in the 1980s.

For the entire 10 year period, the null hypothesis is rejected at the 10 % significance level, and for the first 5 years it is rejected at the 5 % significance level. Conversely, the overall tests of the two other 10 year periods support the CAPM. As mentioned, most empirical tests of the model were conducted before the 1990s. Interestingly, the model seems to have performed a lot better since then, excluding the first 5 year subperiod of the new century. Especially the periods 01/1991-12/1995 and 01/2006-12/2010 seem to favor CAPM.

We can now use $J_7$ to check whether the inference is robust to temporal dependence, heteroskedasticity and non-normality. For the computation of $q$, we have used the formula of Newey and West (1994) provided in section 3.1.5. For the 30 year period we have $q = 4(60/100)^2/9 = 5$ rounded down. For the other intervals, inserting $T = 120$ and $T = 60$ and rounding down yields 4 and 3 respectively. However, it should be noted that $J_7$ is only asymptotically chi-square distributed with $N$ degrees of freedom. Thus, it may tend to reject too often in small samples. Fortunately, MacKinlay and Richardson (1991) show that the small sample
behavior of the Wald test $J_0$ is almost identical to that of the GMM test statistic $J_7$. Also, they suggest comparing the standard Wald test with the exact $F$ test to check whether the use of large sample theory is a problem. In the 5 year subperiods $J_7$ tend to reject more often than $J_1$. However, as $J_0$ produces the same results as $J_7$ the rejections are probably more due to its perverse small sample behavior than violations of the distributional assumptions. Conversely, in the second 10 year subperiod, the asymptotic Wald test agree with the exact $F$-test in its failure to reject $H_0$ at the 5% significance level. Here, we should probably trust the rejection of $H_0$ using $J_7$, and infer that the distributional assumptions are violated. For the entire period, the robust test also cannot reject the efficiency of the CRSP value-weighted proxy.

All in all, the tests all fail to produce evidence against the CAPM for the period as a whole. However, as indicated, the parameter estimates may be biased when using a relatively long time horizon. Unfortunately, the inferences drawn from the subsamples are somewhat inconclusive. In some periods, the model seems to hold, in others, it does not. To check whether the type of proxy used has influence on the conclusions, the test statistics and $p$-values are recalculated using the CRSP equal-weight proxy of the market portfolio.

**Table 4.5: Tests statistic values and $p$-values for the tests of null-hypothesis $\alpha = 0$ using the CRSP equal-weight proxy for the market portfolio.**

<table>
<thead>
<tr>
<th>Time</th>
<th>$J_0$</th>
<th>$p$-value</th>
<th>$J_1$</th>
<th>$p$-value</th>
<th>$J_2$</th>
<th>$p$-value</th>
<th>$J_3$</th>
<th>$p$-value</th>
<th>$J_7$</th>
<th>$p$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>30 year period</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>01/1981-12/2010</td>
<td>3.492</td>
<td>0.967</td>
<td>0.339</td>
<td>0.970</td>
<td>3.475</td>
<td>0.968</td>
<td>3.407</td>
<td>0.970</td>
<td>3.170</td>
<td>0.977</td>
</tr>
<tr>
<td>10 year subperiods</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>01/1981-12/1990</td>
<td>14.263</td>
<td>0.161</td>
<td>1.296</td>
<td>0.242</td>
<td>13.477</td>
<td>0.198</td>
<td>12.691</td>
<td>0.241</td>
<td>16.935</td>
<td>0.076</td>
</tr>
<tr>
<td>01/1991-12/2000</td>
<td>7.656</td>
<td>0.662</td>
<td>0.695</td>
<td>0.727</td>
<td>7.422</td>
<td>0.685</td>
<td>6.989</td>
<td>0.727</td>
<td>11.114</td>
<td>0.349</td>
</tr>
<tr>
<td>01/2001-12/2010</td>
<td>12.739</td>
<td>0.239</td>
<td>1.157</td>
<td>0.328</td>
<td>12.107</td>
<td>0.278</td>
<td>11.401</td>
<td>0.327</td>
<td>15.728</td>
<td>0.108</td>
</tr>
<tr>
<td>5 year subperiods</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>01/1981-12/1985</td>
<td>6.668</td>
<td>0.756</td>
<td>0.545</td>
<td>0.850</td>
<td>6.323</td>
<td>0.787</td>
<td>5.585</td>
<td>0.849</td>
<td>9.016</td>
<td>0.531</td>
</tr>
<tr>
<td>01/1986-12/1990</td>
<td>42.906</td>
<td>0.000</td>
<td>3.504</td>
<td>0.002</td>
<td>32.368</td>
<td>0.000</td>
<td>28.592</td>
<td>0.001</td>
<td>47.365</td>
<td>0.000</td>
</tr>
<tr>
<td>01/1991-12/1995</td>
<td>15.708</td>
<td>0.108</td>
<td>1.283</td>
<td>0.266</td>
<td>13.952</td>
<td>0.175</td>
<td>12.324</td>
<td>0.264</td>
<td>15.567</td>
<td>0.113</td>
</tr>
<tr>
<td>01/1996-12/2000</td>
<td>16.062</td>
<td>0.098</td>
<td>1.312</td>
<td>0.251</td>
<td>14.232</td>
<td>0.163</td>
<td>12.572</td>
<td>0.249</td>
<td>20.301</td>
<td>0.027</td>
</tr>
<tr>
<td>01/2001-12/2005</td>
<td>23.296</td>
<td>0.010</td>
<td>1.903</td>
<td>0.067</td>
<td>19.683</td>
<td>0.032</td>
<td>17.387</td>
<td>0.066</td>
<td>38.359</td>
<td>0.000</td>
</tr>
<tr>
<td>01/2006-12/2010</td>
<td>11.721</td>
<td>0.304</td>
<td>0.957</td>
<td>0.492</td>
<td>10.707</td>
<td>0.381</td>
<td>9.458</td>
<td>0.489</td>
<td>17.816</td>
<td>0.058</td>
</tr>
</tbody>
</table>

Note: For $J_7$ we use $q = 5$ for the 30 year period, $q = 4$ for the 10 year subperiods, and $q = 3$ for the 5 year subperiods.

According to table 4.5, switching proxy seems to have little effect on the overall inference. The change seems to favor the CAPM though, as the $p$-values on the entire period are extremely high now. This may be due to the higher ex-post Sharpe ratio of 0.129 of the equal-weighted portfolio vs. 0.116 of the value-weight portfolio. Recall from the result of Gibbons et al. (1989) that a higher Sharpe ratio leads to a smaller alpha, all else being equal. The test statistics of the 10 year subperiods are now all insignificant. Again, some of the 5 year subperiods favor the CAPM, others do not.

In terms of evidence for or against the CAPM, the picture remains somewhat unclear when evaluating the mean-variance efficiency of the proxies for the market portfolio. Luckily, there are other testable implications. We now turn to the cross-sectional analysis.
### 4.4.3 Cross-sectional tests

In the cross-sectional framework, we regress excess returns on the estimated betas from above. The results are estimates for the intercept and slope of the empirical relation between expected excess return and beta. As before, CAPM would predict an intercept of zero. We had $H_0: \gamma_0 = 0$ against $H_A: \gamma_0 \neq 0$. Furthermore, CAPM predicts that investors are rewarded for bearing covariance risk, so the market risk premium, or equivalently the slope, should be positive. We had $H_0: \gamma_1 = 0$ against $H_A: \gamma_1 \neq 0$. One could be tempted to gain power by conducting a one-sided test where the alternative is $\gamma_1 > 0$. However, we do not want to “miss” an effect in the opposite direction. Especially not when there are several previous tests indicating a somewhat flat empirical relationship. The estimates and test results are presented in table 4.6.

**Table 4.6: Tests statistic values and p-values for the tests of the cross-sectional model using the CRSP value-weight proxy for the market portfolio.**

<table>
<thead>
<tr>
<th>Time</th>
<th>$\hat{\gamma}_0$</th>
<th>$\hat{\sigma}_{\gamma_0}$</th>
<th>$\omega(\hat{\gamma}_0)$</th>
<th>$p$-value</th>
<th>$\hat{\gamma}_1$</th>
<th>$\hat{\sigma}_{\gamma_1}$</th>
<th>$\omega(\hat{\gamma}_1)$</th>
<th>$p$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>30 year period</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>01/1981-12/2010</td>
<td>-0.002</td>
<td>0.008</td>
<td>-0.322</td>
<td>0.748</td>
<td>0.008</td>
<td>0.008</td>
<td>1.038</td>
<td>0.300</td>
</tr>
<tr>
<td>10 year subperiods</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>01/1981-12/1990</td>
<td>0.000</td>
<td>0.010</td>
<td>0.003</td>
<td>0.998</td>
<td>0.003</td>
<td>0.011</td>
<td>0.272</td>
<td>0.786</td>
</tr>
<tr>
<td>01/1991-12/2000</td>
<td>0.009</td>
<td>0.008</td>
<td>1.183</td>
<td>0.239</td>
<td>0.001</td>
<td>0.009</td>
<td>0.116</td>
<td>0.908</td>
</tr>
<tr>
<td>01/2001-12/2010</td>
<td>-0.023</td>
<td>0.009</td>
<td>-2.496</td>
<td>0.014</td>
<td>0.026</td>
<td>0.010</td>
<td>2.498</td>
<td>0.014</td>
</tr>
<tr>
<td>5 year subperiods</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>01/1981-12/1985</td>
<td>-0.005</td>
<td>0.012</td>
<td>-0.431</td>
<td>0.668</td>
<td>0.010</td>
<td>0.013</td>
<td>0.723</td>
<td>0.472</td>
</tr>
<tr>
<td>01/1986-12/1990</td>
<td>-0.015</td>
<td>0.014</td>
<td>-1.072</td>
<td>0.288</td>
<td>0.015</td>
<td>0.016</td>
<td>0.961</td>
<td>0.340</td>
</tr>
<tr>
<td>01/1991-12/1995</td>
<td>0.018</td>
<td>0.010</td>
<td>1.757</td>
<td>0.084</td>
<td>-0.005</td>
<td>0.010</td>
<td>-0.490</td>
<td>0.626</td>
</tr>
<tr>
<td>01/1996-12/2000</td>
<td>0.004</td>
<td>0.012</td>
<td>0.309</td>
<td>0.758</td>
<td>0.004</td>
<td>0.014</td>
<td>0.263</td>
<td>0.793</td>
</tr>
<tr>
<td>01/2001-12/2005</td>
<td>-0.031</td>
<td>0.013</td>
<td>-2.320</td>
<td>0.024</td>
<td>0.034</td>
<td>0.015</td>
<td>2.311</td>
<td>0.024</td>
</tr>
<tr>
<td>01/2006-12/2010</td>
<td>-0.008</td>
<td>0.011</td>
<td>-0.762</td>
<td>0.449</td>
<td>0.011</td>
<td>0.013</td>
<td>0.880</td>
<td>0.382</td>
</tr>
</tbody>
</table>

Note: We have used two-sided tests for both hypotheses.

Again, the null hypothesis that the intercept is zero cannot be rejected at the normal significance levels. For the overall period, the $p$-value of the $t$-test is 0.748. Again the period 01/2001-12/2005 seems to stand out, as the $H_0$ regarding the intercept cannot be rejected in this period at the 5% level. The 1980’s do not seem to provide evidence against the CAPM here though.

Evaluating the slope-estimates, we consistently fail to reject the null hypothesis. This clearly contradicts the spirit of the CAPM, as investors should be rewarded for taking on beta risk. However, we cannot reject that the slope equals zero. Thus, we find no relationship between excess return and beta risk. This is consistent with the findings of Fama and French (1992). The implication is that investors only seem to receive the risk free rate regardless of how responsive their portfolio is to market conditions. Only in the subperiod 01/2001-12/2005 (and the corresponding 10 year period) this relationship is rejected. Here, it should be relevant to further specify the implications of the CAPM. The model does not only say that the risk premium should be positive. It explicitly restricts that investors are rewarded the expected excess return on the market portfolio. Therefore, it should be interesting to test whether the observed slope is significantly different from the mean return on the proxy we use. The average excess return on the CRSP value-weighted portfolio in the 5 year period was only 0.00134. The new and more precise
null hypothesis is $H_0: \gamma_1 = 0.00134$. Along the lines of Black et al. (1972) we obtain the $t$-statistic $(\hat{\gamma}_1 - \gamma_1)/\hat{\sigma}_{\hat{\gamma}_1} = (0.034 - 0.00134)/0.015 = 2.2$. The corresponding $p$-value in a Student $t$ distribution with 60-1 degrees of freedom is app. 0.03. This indicates that even though the observed slope was positive in this subperiod, we can reject that it equaled the theoretical correct market risk premium.

To check whether the inferences drawn are sensitive to the portfolio used, the results are recalculated in table 4.7 using the CRSP equal-weight proxy. Again, we generally fail to reject that the intercepts and slopes are zero. Note that in the subperiod 01/1986-12/1990 there is a significant negative slope. This also challenges the CAPM. On the other hand, when evaluating the significant slope estimate of 0.026 from 01/2001-12/2005, one cannot reject that it equals the market risk premium of 0.015. Nonetheless, the overall conclusion prevails: we find no systematic relationship between expected excess return and beta. It should be mentioned that these $t$-tests also rely on the assumption that returns are normally distributed and temporally IID.

<table>
<thead>
<tr>
<th>Time</th>
<th>$\hat{\gamma}_0$</th>
<th>$\hat{\sigma}_{\hat{\gamma}_0}$</th>
<th>$\omega(\hat{\gamma}_0)$</th>
<th>$p$-value</th>
<th>$\hat{\gamma}_1$</th>
<th>$\hat{\sigma}_{\hat{\gamma}_1}$</th>
<th>$\omega(\hat{\gamma}_1)$</th>
<th>$p$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>30 year period</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>01/1981-12/2010</td>
<td>0.004</td>
<td>0.004</td>
<td>0.949</td>
<td>0.343</td>
<td>0.004</td>
<td>0.005</td>
<td>0.777</td>
<td>0.438</td>
</tr>
<tr>
<td><strong>10 year subperiods</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>01/1981-12/1990</td>
<td>0.016</td>
<td>0.008</td>
<td>2.035</td>
<td>0.044</td>
<td>-0.013</td>
<td>0.009</td>
<td>-1.527</td>
<td>0.129</td>
</tr>
<tr>
<td>01/1991-12/2000</td>
<td>0.010</td>
<td>0.005</td>
<td>1.844</td>
<td>0.068</td>
<td>0.000</td>
<td>0.007</td>
<td>0.047</td>
<td>0.962</td>
</tr>
<tr>
<td>01/2001-12/2010</td>
<td>-0.012</td>
<td>0.006</td>
<td>-2.070</td>
<td>0.041</td>
<td>0.020</td>
<td>0.008</td>
<td>2.539</td>
<td>0.012</td>
</tr>
<tr>
<td><strong>5 year subperiods</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>01/1981-12/1985</td>
<td>0.002</td>
<td>0.008</td>
<td>0.251</td>
<td>0.802</td>
<td>0.004</td>
<td>0.010</td>
<td>0.393</td>
<td>0.696</td>
</tr>
<tr>
<td>01/1986-12/1990</td>
<td>0.036</td>
<td>0.014</td>
<td>2.621</td>
<td>0.011</td>
<td>-0.035</td>
<td>0.015</td>
<td>-2.369</td>
<td>0.021</td>
</tr>
<tr>
<td>01/1991-12/1995</td>
<td>0.007</td>
<td>0.005</td>
<td>1.311</td>
<td>0.195</td>
<td>0.008</td>
<td>0.007</td>
<td>1.153</td>
<td>0.253</td>
</tr>
<tr>
<td>01/1996-12/2000</td>
<td>0.015</td>
<td>0.009</td>
<td>1.559</td>
<td>0.124</td>
<td>-0.008</td>
<td>0.012</td>
<td>-0.658</td>
<td>0.513</td>
</tr>
<tr>
<td>01/2001-12/2005</td>
<td>-0.015</td>
<td>0.008</td>
<td>-1.949</td>
<td>0.056</td>
<td>0.026</td>
<td>0.011</td>
<td>2.497</td>
<td>0.015</td>
</tr>
<tr>
<td>01/2006-12/2010</td>
<td>-0.005</td>
<td>0.008</td>
<td>-0.588</td>
<td>0.559</td>
<td>0.010</td>
<td>0.012</td>
<td>0.869</td>
<td>0.389</td>
</tr>
</tbody>
</table>

Note: We have used two-sided tests for both hypotheses.

**4.4.4 Graphical interpretation and conclusion**

Figure 4.1 provides a geometrical illustration of the parameter estimates from the 30 year period and 10 year subperiods. Also, the theoretical correct relationship from the CAPM between excess return and beta is depicted as a red line. This is the excess return version of the SML and its equation is found under each graph. It intercepts the vertical axis at zero and has a slope equal to the average excess return on the market portfolio proxy. The exact linear relationship in the CAPM implies that there is no dispersion around the SML. In section 4.4.2 this implication translates into
the hypothesis $H_0: \alpha = 0$. To see why note that the alpha estimates$^{17}$ are the vertical distances between the observations (blue squares) and the red SML line in figure 4.1. Thus, testing the exact linear relationship is equivalent to testing these distances against zero simultaneously. From Figure 4.1A, it should be clear why we failed to reject the null hypothesis for the overall period. The observations are practically on the SML.

The estimated relationship from the cross-sectional regression is also depicted in figure 4.1. It is the black, thin line that intercepts the vertical axis at $\hat{\beta}_0$ and has a slope of $\hat{\beta}_1$. If the CAPM holds, this line should be identical to the red SML line. In the cross-sectional regression we tested if the intercept was significantly different from zero. A zero intercept implies that investors receive exactly the risk free rate when they bear no covariance risk. This is consistent with the Sharpe-Lintner CAPM. As for the slope, we started by testing the observed coefficient on beta against zero. If it was significantly different, we tested whether it equaled the slope on the SML. Some of the early tests reported that the empirical relationship was flatter than the theoretical counterpart. In 4.1A this is not evident, as the black line is steeper than the SML. However, we could not reject that the slope was zero, so statistically, the flat relationship is confirmed.

In figure 4.2 the empirical relationship of each 5 year subperiod is plotted against the excess return SML. Here the significant positive slope from 01/2001-12/2005 is evident in figure 4.2E. In the graphs 4.2B-E there is an extreme outlier that seems to distort the picture. This is portfolio 10 containing the smallest companies in the sample. They seem to be too noisy for the CAPM to have success predicting their excess returns.

Figure 4.2B illustrates the efficiency problems inherent in working with small samples. In the period 01/1986-12/1990 the slope of the empirical relationship seems to be rather steep. This is confirmed in table 4.6 where the mean monthly excess return is 0.015. This corresponds to an annualized, empirical market risk premium of 18% which in economic terms seems to be very significant. However, the monthly standard deviation on this estimate of 0.016 is so high that the null hypothesis cannot be rejected.

Let us summarizes what we have found. When evaluating the mean-variance efficiency of the market portfolio proxy, the results were not entirely consistent. Testing the entire period, we found no evidence against the CAPM. Beta seemed to completely explain excess returns. However, when evaluating the subperiods the picture became unclear, as some intervals seemed to favor the model, where others produced evidence against it. The inferences with an equal-weighted proxy were similar. Turning to cross-sectional analysis, we found clear evidence against the CAPM. Specifically, we could not reject that the slope of the empirical relation between excess returns and beta is zero. Thus, contrary to the predictions of the CAPM, it seemed that investors are not rewarded for bearing covariance risk. For further research, it should be interesting to focus more on the cross-sectional tests of the model. Specifically, one could test the effects documented in the anomalies literature on newer data. It seemed that the model performed rather well in the two last decades (excluding 01/2001-12/2005).

$^{17}$Here alpha is equivalent to the performance measure Jensen’s alpha.
Figure 4.1: Annualized excess returns vs. beta in the 30 year period and the 10 year subperiods using the CRSP value-weighted proxy of the market portfolio.

A: 01/1981-12/2010

\[ E[Z_i] = \beta_{im} \times 6.42\% \]

B: 01/1981-12/1990

\[ E[Z_i] = \beta_{im} \times 4.88\% \]

C: 01/1991-12/2000

\[ E[Z_i] = \beta_{im} \times 11.99\% \]

D: 01/2001-12/2010

\[ E[Z_i] = \beta_{im} \times 2.39\% \]

Note: The horizontal axis of the diagrams begins at \( \beta_{im} = 0.5\), to focus on the observations. Below each diagram the theoretical correct relationship is found. This is the excess return version of the Sharpe-Lintner CAPM model. In each graph it is depicted as a red line – the SML. The thin black line is the relationship estimated in the cross-sectional model.
Figure 4.2: Annualized excess returns vs. beta in the 5 year subperiods using the CRSP value-weighted proxy of the market portfolio.

Note: The horizontal axis of the diagrams begins at 0.5, to focus on the observations. Below each diagram the theoretical correct relationship is found. This is the excess return version of the Sharpe-Lintner CAPM model. In each graph it is depicted as a red line - the SML. The thin black line is the relationship estimated in the cross-sectional model.
Chapter 5 Alternative Asset Pricing Models

In this chapter we shortly introduce other popular models of asset pricing. They address some of the deficiencies of the traditional CAPM. The first is the Black version of the CAPM. As mentioned, this had early empirical success, but has later also been criticized heavily by researchers. After discussing this model, multifactor pricing models will be considered. The anomalies literature mentioned in chapter 4 indicates, that one risk factor does not suffice to explain returns. Thus, we briefly examine one major contribution to the field of multifactor models named the Arbitrage Pricing Theory (APT).

5.1 The Black Version

In 1972 Black derived a more general case of the CAPM model. He removed the risk-free rate and replaced it with a zero-beta portfolio. This portfolio has the minimum variance among portfolios uncorrelated with the market. It can be shown, that in the zero-beta portfolio plots directly below the market portfolio on the inefficient part of the minimum variance frontier from chapter 2 (Campbell et al., 1997). In addition to this change, Black also assumes that unlimited short selling of the risky assets is possible. It has been argued, that this is in fact is just as unrealistic as the existence of a constant risk-free rate. The Black version of the CAPM is given by

\[ E[R_i] = E[R_{om}] + \beta_{lm}(E[R_m] - E[R_{om})], \]  

(5.1)

where \( E[R_{om}] \) is the expected return on the zero-beta portfolio. Note that the traditional CAPM emerges if we assume that the return on the zero-beta portfolio equals the risk-free rate. Conversely, Black just restricts that \( E[R_{om}] \) must be below the expected market return. That is, it is still assumed that investors are rewarded for taking on covariance risk. As mentioned, the Black version had some early empirical success. It is apparent in (5.1) why it could explain the nonzero alphas observed when testing the traditional model. Strictly, the Black model only predicts that beta suffices to explain expected returns, and that the coefficient on beta is positive (Fama, 2004). Testing the Black version is more complicated than testing the original model, as the econometric techniques treat the zero-beta portfolio as unobserved (Campbell et al., 1997).

5.2 Arbitrage Pricing Theory

Ross (1976) launched APT as an alternative to the Sharpe-Lintner model. Unlike the CAPM, it permits more than one risk factor. This may be appropriate, given the vast amount of literature documenting that beta does not completely explain expected returns\(^\text{18}\). Essentially, it is based on the law of one price, stating that two equivalent goods should have the same price. If this law is not upheld, arbitrage possibilities exist. That is, one can make a riskless profit by selling the overpriced goods and investing the proceeds in the underpriced one. APT assumes that there is no arbitrage, and that markets are competitive and frictionless. CAPM was based on the assuming that investors are mean-variance optimizers. Instead, APT builds on the process that generates security returns. It assumes that asset returns are linearly related to a set of factors or indices. Also, there is a set of factor loadings or sensitivities measuring the assets’ responsiveness to the factors. In the CAPM framework the single factor loading would be \( \beta_{lm} \). Assume that the return generating process for the \( N \) asset case with \( K \) factors is

\[ R = a + Bf + \varepsilon, \]  

(5.2)

\(^{18}\) See chapter 4 for references.
where

$$E[\varepsilon | \mathbf{f}] = 0$$

and

$$E[\varepsilon \varepsilon' | \mathbf{f}] = \Sigma.$$  

Here $\mathbf{R}$ is a $N \times 1$-vector of asset returns, $\mathbf{a}$ is a $N \times 1$-vector of intercepts, $\mathbf{B}$ is a $N \times K$-matrix of factor loadings, $\mathbf{f}$ is $K \times 1$-vector of factors, and $\varepsilon$ is a $N \times 1$-vector of disturbances. The $K$ factors explain the common variation in returns, so the idiosyncratic or specific risk, denoted by the disturbances $\varepsilon$, disappears in properly diversified portfolios. Given that there is no arbitrage, it can be shown from (5.2) that the following approximation applies

$$\mu \approx \mathbf{a} + \mathbf{B} \lambda_K,$$  

(5.3)

where $\mu$ a $N \times 1$-vector of expected returns, $\mathbf{a}$ is a $N \times 1$-vector of ones, $\lambda_0$ is the zero-beta return (preferably a riskless rate), and $\lambda_K$ is the $K \times 1$-vector of risk premia associated with the factors in (5.2). As (5.3) is only an approximation, we can neither test nor reject the model, unless arbitrage possibilities are discovered. Thus, additional assumptions has to be imposed to obtain an exact relation. On the other hand, it has been documented that deviations from exact pricing are insignificant, so that we can assume an exact relation in imperical analysis anyway. Obviously (5.3) looks a lot like the CAPM relation we have focus on in the previous chapters. Interestingly it can be shown, that CAPM is in fact just a special case of this model (Campbell et al., 1997; Elton and Gruber, 1995).
Chapter 6 Conclusion

The thesis has sought answers to the three research questions regarding The Capital Asset Pricing Model (CAPM) listed in chapter 1. In answering these three questions, the thesis was divided into three main sections. The first dealt with the theoretical fundament of the model, the next dealt with the statistical framework for testing it, and the third was an empirical examination. Also, a brief introduction to other popular models of asset pricing was given.

The CAPM has its roots in mean-variance analysis (MVA) which was introduced by Markowitz in 1952. MVA is a framework for selecting efficient portfolios based on the assumption that investors like wealth and dislike risk (i.e. are risk averse). An efficient portfolio has the maximum expected return of all portfolios return for a given variance. MVA can be theoretically motivated by assuming quadratic utility functions or that returns follow a distribution from the class of elliptical distributions.

When a risk-free asset exists, it can be shown that all efficient portfolios are a combination of a basket of risk assets and the risk-free asset. Assuming concave utility functions, all investors will choose such a combination. Interestingly, with homogenous expectations and an exogenous risk-free rate, they will all use the same basket of risk assets regardless of their individual preferences. This result is known as the separation theorem. If everyone holds some proportion of the same portfolio of risky assets, this portfolio must be the market portfolio. The efficient portfolios combining the market portfolio and the risk-free rate are all placed on what is known as the Capital Market Line (CML) in a mean-standard deviation space.

The CML is not a sufficient framework though, as the standard deviation only explains the expected return on efficient portfolios. A quantitative risk measure that works for any feasible asset should be something that measures the asset’s contribution to the total risk of a portfolio. It can be shown that beta can fulfill this role. It is defined as the covariance of returns between the asset and the market portfolio divided by the variance of the market portfolio. By using beta, we get an equation relating the risk of any asset to its expected return. This is the CAPM model. Intuitively, beta as a risk measure makes sense, as rational investors will hold diversified portfolios. Thus, they are only exposed to systematic risk and should therefore only be rewarded for their portfolio’s sensitivity to the overall economic activity.

In the mathematical derivations of the CAPM, it is evident that the exact linear relationship in the model only holds when the market portfolio is efficient. Thus, a central, testable implication of the CAPM is the efficiency of the market portfolio. In empirical applications, this is equivalent to testing the intercept against zero in the excess return version of the traditional model. Testing $N$ assets, the null hypothesis is that all the estimated $N$ intercepts are jointly zero. Several statistics allow for joint tests. A number of them are only asymptotically distributed under the null-hypothesis. One such test statistic is a Wald type test. Another is the likelihood ratio test. In finite samples, these asymptotic approximations can behave rather perversely. They tend to reject too often. Fortunately, the Wald type test can be transformed into an exact $F$ test, and the likelihood ratio test can be corrected to perform much better in finite samples. All the above test statistics rely on the assumption that returns are independently, identically distributed (IID) and jointly multivariate normal. There is a vast amount of literature indicating that this assumption is too strong. Luckily, a robust Wald type test can be derived using the Generalized Method of Moments.
(GMM) approach. This can serve as a robustness check of the inferences drawn from the above test statistics.

There are also other testable implications of the CAPM than the mean-variance efficiency of the market portfolio. One is that the market risk premium is positive. Another is that no other risk measure has explanatory power regarding expected returns, once beta has been accounted for. These implications can be tested cross-sectionally using simple $t$ tests. The cross-sectional approach allows for more in-depth tests of other risk measures. Thus, one can check whether beta completely explains the cross-sectional variation in expected returns when other firm characteristics are added to the model. If it does, the coefficient on the other risk measure should be zero.

When implementing these tests, there are a number of issues to consider. A very important one is the point that one can really never test the CAPM, as it involves the market portfolio of all risky assets. Thus, we actually only test the mean-variance efficiency of the market proxy. Luckily, it has been documented that a proxy which is highly correlated with the true market portfolio suffices, and that inferences are not sensitive to the inclusion of the likes of real-estate and bonds. Another point to consider is the tradeoff between non-stationarity of the model's parameter estimates and statistical power. This is relevant in choosing the right period length when drawing samples. The most commonly used is 5 years of monthly data.

We have tested the CAPM on a sample of American stocks assigned to 10 size-sorted portfolios. Monthly, total returns were collected from January 1980 to December 2010. Regarding the mean-variance efficiency of the proxy, there is no definitive conclusion. If we are to trust the overall test statistics, the result is unambiguous. We cannot reject the null-hypothesis, so beta seems to completely explain expected returns. However, the long period estimates can be biased as the true parameter values are likely to change over time. Unfortunately, the results from the subperiods of 5 years were somewhat inconsistent. However, excluding the years 2001-2005, the data after the 1990s seems to fit the model somewhat better than previous periods do. This suggests that a renewed test of documented anomalies such as the size-effect may be appropriate.

Testing the slope of the empirical relation between beta and excess returns, we did not find a significant, positive risk premium. Thus, the test indicated that there is no systematic relationship between beta and excess returns. This is clearly evidence against the CAPM. It indicates that investors are not rewarded for bearing covariance risk which obviously contradicts the theoretical anchor of the model.

Among alternative models, we have the Black version. This version of the CAPM does not assume anything about a risk-free rate. The model had early empirical success, as it could explain the deviations from the exact linear relationship expressed in the traditional version. Another framework is the Arbitrage Pricing Theory (APT). This allows for several risk factors which empirically can be motivated by the anomalies literature rejecting beta as a solemn risk measure.
Bibliography


Appendix

Appendix A.1

The vectors and the matrix in the variance term are defined as:

\[
\omega' = \begin{bmatrix} X_1 & X_2 & \cdots & X_N \end{bmatrix},
\]
\[
\Omega = \begin{bmatrix}
\sigma_{11} & \sigma_{12} & \cdots & \sigma_{1N} \\
\sigma_{21} & \sigma_{22} & \cdots & \sigma_{2N} \\
\vdots & \vdots & \ddots & \vdots \\
\sigma_{N1} & \sigma_{N2} & \cdots & \sigma_{NN}
\end{bmatrix},
\]
\[
\omega = \begin{bmatrix} X_1 \\
X_2 \\
\vdots \\
X_N
\end{bmatrix}.
\]

In \( \Omega \) the diagonal composes the variances and the rest is covariances. As \( \omega' \) is a \( 1 \times N \)-vector and \( \Omega \) is a \( N \times N \)-matrix, their product exists, and it is a \( 1 \times N \)-vector:

\[
\omega' \Omega = \begin{bmatrix} X_1 & X_2 & \cdots & X_N \end{bmatrix} \begin{bmatrix}
\sigma_{11} & \sigma_{12} & \cdots & \sigma_{1N} \\
\sigma_{21} & \sigma_{22} & \cdots & \sigma_{2N} \\
\vdots & \vdots & \ddots & \vdots \\
\sigma_{N1} & \sigma_{N2} & \cdots & \sigma_{NN}
\end{bmatrix} = \begin{bmatrix} X_1(X_1 \sigma_{11} + X_2 \sigma_{21} + \cdots + X_N \sigma_{N1}) + X_2(X_1 \sigma_{12} + X_2 \sigma_{22} + \cdots + X_N \sigma_{N2}) + \cdots + X_N(X_1 \sigma_{1N} + X_2 \sigma_{2N} + \cdots + X_N \sigma_{NN}) \end{bmatrix}.
\]

As \( (\omega' \Omega) \omega \) is a \( 1 \times N \)-vector and \( \omega \) is a \( N \times 1 \)-column vector their product exists and is a \( 1 \times 1 \)-matrix also known as a scalar:

\[
(\omega' \Omega) \omega = X_1(X_1 \sigma_{11} + X_2 \sigma_{21} + \cdots + X_N \sigma_{N1}) + X_2(X_1 \sigma_{12} + X_2 \sigma_{22} + \cdots + X_N \sigma_{N2}) + \cdots + X_N(X_1 \sigma_{1N} + X_2 \sigma_{2N} + \cdots + X_N \sigma_{NN}).
\]

There are \( N \) terms in (A.1.5) containing the variances given by

\[
X_1X_1 \sigma_{11} + X_2X_2 \sigma_{22} + \cdots + X_NX_N \sigma_{NN} = X_1^2 \sigma_1^2 + X_2^2 \sigma_2^2 + \cdots + X_N^2 \sigma_N^2.
\]

This is the first part in the formula for the portfolio variance using summation notation. The second part contains the covariances. Thus, the much more elegant expression \( \omega' \Omega \omega \) equals the formula

\[
\sum_{j=1}^{N} (X_j^2 \sigma_j^2) + \sum_{j=1}^{N} \sum_{k=1}^{N} (X_jX_k \sigma_{jk}).
\]

\(^{19}\)Their products exists, because the number of columns in the first vector equals the number of rows in the second. The result is a vector with the same amount of rows as the first and the same amount of columns as the second.
Appendix A.2
We had
\[ \mu_p = \delta^1 (\mu'\Omega^{-1}\mu) + \delta^2 (\mu'\Omega^{-1}i) \]  
(A.2.1.a)
and
\[ 1 = \delta^1 (i'\Omega^{-1}\mu) + \delta^2 (i'\Omega^{-1}i). \]  
(A.2.2 a)
Solving (A.2.1 a) for \( \delta^1 \)
\[ \delta^1 = \frac{\mu_p - \delta^2 (\mu'\Omega^{-1}i)}{(\mu'\Omega^{-1}\mu)}. \]  
(A.2.1 b)
Solving (A.2.2 a) for \( \delta^2 \)
\[ \delta^2 = \frac{1 - \delta^1 (i'\Omega^{-1}\mu)}{(i'\Omega^{-1}i)}. \]  
(A.2.2 b)
Inserting (A.2.2 b) in (A.2.1 b) we can write
\[ \delta^1 (\mu'\Omega^{-1}\mu) = \mu_p - \left[ \frac{1 - \delta^1 (i'\Omega^{-1}\mu)}{(i'\Omega^{-1}i)} \right] (\mu'\Omega^{-1}i) \]
\[ \iff \]
\[ \delta^1 (\mu'\Omega^{-1}\mu)(i'\Omega^{-1}i) = (i'\Omega^{-1}i)\mu_p - [(i'\Omega^{-1}\mu) - \delta^1 (i'\Omega^{-1}\mu)^2] \]
\[ \iff \]
\[ \delta^1 (\mu'\Omega^{-1}\mu)(i'\Omega^{-1}i) - \delta^1 (i'\Omega^{-1}\mu)^2 = (i'\Omega^{-1}i)\mu_p - (i'\Omega^{-1}\mu) \]
\[ \iff \]
\[ \delta^1 = \frac{(i'\Omega^{-1}i)\mu_p - (i'\Omega^{-1}i)}{(\mu'\Omega^{-1}\mu)(i'\Omega^{-1}i) - (i'\Omega^{-1}\mu)^2} \]
\[ = \frac{C\mu_p - A}{D}, \]  
(A.2.3)
where \( A = i'\Omega^{-1}\mu, B = \mu'\Omega^{-1}\mu, C = i'\Omega^{-1}i \) and \( D = BC - A^2 \). In the second line we used the fact that \((i'\Omega^{-1}\mu) = (\mu'\Omega^{-1}i)\). Now let us derive the expression for \( \delta^2 \). First substitute (A.2.1 b) into (A.2.2 b). We can then write
\[ \delta^2 (i'\Omega^{-1}i) = 1 - \frac{\mu_p - \delta^2 (\mu'\Omega^{-1}i)}{(\mu'\Omega^{-1}\mu)} (i'\Omega^{-1}\mu) \]
\[ \iff \]
\[ \delta^2 (i'\Omega^{-1}i)(\mu'\Omega^{-1}\mu) = (\mu'\Omega^{-1}\mu) - [\mu_p(\mu'\Omega^{-1}i) - \delta^2 (\mu'\Omega^{-1}i)^2] \]
\[ \iff \]
\[ \delta^2 (i'\Omega^{-1}i)(\mu'\Omega^{-1}\mu) - \delta^2 (\mu'\Omega^{-1}i)^2 = (\mu'\Omega^{-1}\mu) - \mu_p (\mu'\Omega^{-1}i) \]
\[ \delta^2 = \frac{(\mu' \Omega^{-1} \mu) - \mu_p (\mu' \Omega^{-1} \mu)}{(i' \Omega^{-1} i)(\mu' \Omega^{-1} \mu) - (\mu' \Omega^{-1} \mu)^2} = \frac{B - A \mu_p}{D}. \]

In the second line we again used the fact that \((i' \Omega^{-1} \mu) = (\mu' \Omega^{-1} \mu)\).

### Appendix A.3

We had

\[ \omega_p = \delta^1 (\Omega^{-1} \mu) + \delta^2 (\Omega^{-1} i) \]  

and

\[ \delta^1 = \frac{C \mu_p - A}{D}, \]

\[ \delta^2 = \frac{B - A \mu_p}{D}. \]

Substituting the expressions for the lagrange multipliers into (A.3.1) we get

\[ \omega_p = \frac{C \mu_p - A}{D} (\Omega^{-1} \mu) + \frac{B - A \mu_p}{D} (\Omega^{-1} i) \]

\[ = \frac{1}{D} \left[ C (\Omega^{-1} \mu) \mu_p - A (\Omega^{-1} \mu) \right] + \frac{1}{D} \left[ B (\Omega^{-1} i) - A \mu_p (\Omega^{-1} i) \right] \]

\[ = \frac{1}{D} \left[ B (\Omega^{-1} i) - A (\Omega^{-1} \mu) + C (\Omega^{-1} \mu) \mu_p - A \mu_p (\Omega^{-1} i) \right] \]

\[ = \frac{1}{D} \left[ B (\Omega^{-1} i) - A (\Omega^{-1} \mu) \right] + \frac{1}{D} \left[ C (\Omega^{-1} \mu) - A (\Omega^{-1} i) \right] \mu_p \]

\[ = g + h \mu_p, \]  

where

\[ g = \frac{1}{D} \left[ B (\Omega^{-1} i) - A (\Omega^{-1} \mu) \right] \]

and

\[ h = \frac{1}{D} \left[ C (\Omega^{-1} \mu) - A (\Omega^{-1} i) \right]. \]
Appendix B.1

We had that

$$\text{Var}(\hat{\alpha}) = \text{Var} \left[ \frac{1}{T} \sum_{t=1}^{T} (Z_t) \right] + \text{Var} \left[ \frac{1}{T} \sum_{t=1}^{T} (Z_t - \hat{\mu})(Z_{mt} - \hat{\mu}_m) \right] \quad (B.1.1)$$

Let us consider one term at a time. For the first term we use the rule \( \text{Var}(aX) = a^2 \text{Var}(X) \) and the rule \( \text{Var}(\sum_{t=1}^{T} X_t) = \sum_{t=1}^{T} \text{Var}(X_t) \), where the \( X_t \)’s are uncorrelated, recalling that excess returns are assumed to be temporally independent. Thus we have

$$\text{Var} \left[ \frac{1}{T} \sum_{t=1}^{T} (Z_t) \right] = \frac{1}{T^2} \sum_{t=1}^{T} \text{Var}(Z_t)$$

$$= \frac{1}{T} \sum_{t=1}^{T} \Sigma$$

$$= \frac{1}{T} \Sigma. \quad (B.1.2)$$

The second equality follows from the fact that the variance of the dependent variable in the market model equals the variance of the error term. In the second part of (B.1.1) note that

$$\sum_{t=1}^{T} (Z_t - \hat{\mu})(Z_{mt} - \hat{\mu}_m)$$

$$= \sum_{t=1}^{T} \left( Z_t - \frac{1}{T} \sum_{t=1}^{T} (Z_t) \right) \left( Z_{mt} - \frac{1}{T} \sum_{t=1}^{T} (Z_{mt}) \right)$$

$$= \sum_{t=1}^{T} Z_t Z_{mt} - \sum_{t=1}^{T} Z_t \left[ \frac{1}{T} \sum_{t=1}^{T} (Z_{mt}) \right] - \sum_{t=1}^{T} \left[ \frac{1}{T} \sum_{t=1}^{T} (Z_t) \right] Z_{mt}$$

$$+ T \left[ \frac{1}{T} \sum_{t=1}^{T} (Z_t) \right] \left[ \frac{1}{T} \sum_{t=1}^{T} (Z_{mt}) \right]$$

$$= \sum_{t=1}^{T} Z_t Z_{mt} - \frac{1}{T} \sum_{t=1}^{T} \sum_{t=1}^{T} (Z_t) Z_{mt}$$

$$= \sum_{t=1}^{T} Z_t Z_{mt} - \sum_{t=1}^{T} (Z_t) \hat{\mu}_m$$

$$= \sum_{t=1}^{T} Z_t (Z_{mt} - \hat{\mu}_m). \quad (B.1.3)$$

The second equality comes from substituting the definitions of the means into the expression. Inserting (B.1.3) into the second part of (B.1.1) we have

$$\text{Var} \left[ \frac{\sum_{t=1}^{T} (Z_t - \hat{\mu})(Z_{mt} - \hat{\mu}_m)}{\sum_{t=1}^{T} (Z_{mt} - \hat{\mu}_m)^2} \hat{\mu}_m \right]$$
\[
\begin{align*}
&= \text{Var} \left[ \frac{\sum_{t=1}^{T} Z_t (Z_{mt} - \mu_m)}{\sum_{t=1}^{T} (Z_{mt} - \mu_m)^2} \right] \\
&= \sum_{t=1}^{T} \text{Var} \left[ Z_t \frac{(Z_{mt} - \mu_m)}{\sum_{t=1}^{T} (Z_{mt} - \mu_m)^2} \hat{\mu}_m \right] \\
&= \sum_{t=1}^{T} \text{Var} \left[ Z_t \frac{(Z_{mt} - \mu_m)}{\sum_{t=1}^{T} (Z_{mt} - \mu_m)^2} \hat{\mu}_m \right]^2 \\
&= \Sigma \left[ \frac{\sum_{t=1}^{T} (Z_{mt} - \mu_m)}{\sum_{t=1}^{T} (Z_{mt} - \mu_m)^2} \hat{\mu}_m \right]^2 \\
&= \frac{1}{T} \Sigma \frac{\mu_m^2}{\sigma^2_m} \\
\end{align*}
\]

where

\[
\hat{\sigma}_m^2 = \frac{1}{T} \sum_{t=1}^{T} (Z_{mt} - \mu_m)^2. 
\]

The second and third equality follow from the rules applied above (Skovmand, 2011 b). Inserting (B.1.2) and (B.1.4) into (B.1.1) we have

\[
\text{Var}(\bar{\alpha}) = \frac{1}{T} \Sigma + \frac{1}{T} \Sigma \frac{\mu_m^2}{\sigma^2_m} \\
= \frac{1}{T} \left[ 1 + \frac{\mu_m^2}{\sigma^2_m} \right] \Sigma. 
\]

To see where the variance of \( \hat{\beta} \) comes from notice the second last line of the derivation of (B.1.4). If the last \( \mu_m \) is removed we get

\[
\text{Var}(\hat{\beta}) = \frac{1}{T} \left[ \frac{1}{\sigma^2_m} \right] \Sigma. 
\]