

https://dspace.org/

Higher moment asset pricing on the JSE

Bester, Johan

2016
http://hdl.handle.net/10539/20495

Downloaded from a DSpace repository. https://dspace.org/

FINANCE DIVISION,

SCHOOL OF ECONOMIC AND BUSINESS SCIENCES,

FACULTY OF COMMERCE, LAW AND MANAGEMENT

HIGHER MOMENT ASSET PRICING ON THE JSE

M.Com Business Finance

(50% weighting)

JOHAN BESTER

Student Number 881281

SUPERVISORS: Prof CHRISTO AURET and Mr DANIEL PAGE

DATE OF SUBMISSION: 28 APRIL 2015

WORD COUNT: 39 858

ABSTRACT

The purpose of the study is to investigate the effects of relaxing the assumption of multivariate normality typically utilised within the traditional asset pricing framework. This is achieved in two ways. The first involves the introduction of higher moments into the linear Capital Asset Pricing Model while the second involves a Monte Carlo experiment to determine the impact of skewness and kurtosis on test statistics traditionally employed to assess the validity of asset pricing models. We commence by establishing non-normality for the majority of sample portfolios. A cross-sectional regression approach is employed to estimate factor risk premia and test higher moment Capital Asset Pricing Models. Unconditional coskewness and unconditional cokurtosis are found to be priced within the market equity (size) sorted and book equity/market equity (value) sorted portfolio sets over the period January 1993 to December 2013. Conditional coskewness and conditional cokurtosis are found to be priced for only the size sorted portfolios over the period January 1997 to December 2013. Factor risk premia estimated for coskewness are generally positive while risk premia estimated for cokurtosis are negative. This suggests a positive relationship between coskewness and expected return and a negative relationship between cokurtosis and expected return. The results of the asset pricing model tests are mixed. The pricing errors for higher moment Capital Asset Pricing Models are shown to be significantly different from zero for size sorted portfolios while pricing errors on the value sorted, dual size-value sorted and industry portfolios are found to be statistically insignificant. This suggest that none of the asset pricing models tested are the true model as it would explain variation in expected returns regardless of the data generating process. Finally we show that the Ordinary Least Square Wald test statistic has the most desirable size characteristics while the Generalised Least Squares J-test statistic has the most desirable power characteristics when dealing with non-normal data.

Acknowledgements

I would like to thank three people that assisted in the completion of this dissertation. My supervisors, Professor Christo Auret and Mr Daniel Page, for the guidance and support throughout the process. My brother, Landman Bester, for the guidance when I got stuck with the more complex mathematics and programming.

Table of Contents

ABSTRACT ii

Acknowledgements iii

1. Introduction 1

2. Literature Review 5

2.1 Asset Pricing Models 5

2.2 Capital Asset Pricing Model 5

Assumptions 7

The Conditional CAPM 7

Early Tests of the CAPM 8

Anomalies to the CAPM 8

South African Anomalies 10

2.3 Arbitrage Pricing Theory 11

2.4 Fama French Three Factor Model 11

2.5 Higher Moment Capital Asset Pricing Models 15

Skewness in Asset Pricing 16

Kurtosis in Asset Pricing 19

2.6 Estimation and Asset Pricing Tests 20

Ordinary Least Squares 20

Generalised Least Squares 21

Fama-Macbeth Regressions 22

Generalised Method of Moments 23

HLV Test Statistic 25

3. Data 27

4. Methodology 28

4.1 Portfolio Sorts and Factor Construction 28

Size and Value Portfolios 29

Industry Indices 29

FF3F Factors 29

Conditional Skewness and Conditional Kurtosis Factors 30

4.2 Distribution Analysis 30

Individual Portfolios 30

Multivariate Normality 31
Linearity 32
4.3 Estimation 32
Time-series Regressions 32
Cross-sectional Regressions 33
4.4 Asset Pricing Model Tests 33
4.5 Simulations 34
5. Empirical Results 36
5.1 Factor Analysis 36
Size and Value Premia 36
A Coskewness Premium 39
Factor Correlation 41
5.2 Distribution Analysis 41
Value-weighted Portfolio Analysis 41
Equally-weighted Portfolio Analysis 46
Multivariate Normality 50
5.3 Time-Series Regressions 50
Time-series Variation 51
5.4 Cross-sectional Regressions 53
Cross-sectional Variation 54
Conditional vs Unconditional Asset Pricing Models 57
Asset Pricing Model Risk Premia 59
Expected Return, Market Beta, Coskewness Beta and Cokurtosis Beta 65
Asset Pricing Model Tests 67
5.5 Monte Carlo Results 71
6. Discussion and Conclusion 75
References 78
Appendix A: Portfolio and Factor Formation Automation 85
Appendix B: Higher Moment Automation using Visual Basic for Applications 94
Appendix C: Python Code to Establish Class of Regression 97
Appendix D: Python Code for Monte Carlo Experiment 107

List of Tables and Figures

Table 2.1 Extract from Fama and French (1993) results 13
Table 2.2 Extract from Basiewicz and Auret (2010) results 14
Table 3.1 FTSE / JSE industry classification indices 27
Table 5.1 Periodic analysis of factors returns 37
Table 5.2 Factor correlation matrices 40
Table 5.3 Return distribution analysis of value weighted portfolios (Jan 1993 – Dec 2013) 42
Table 5.4 Return distribution analysis of value weighted portfolios (Jan 1997 – Dec 2013) 44
Table 5.5 Return distribution analysis of equally weighted portfolios (Jan 1993 – Dec 2013) 47
Table 5.6 Return distribution analysis of equally weighted portfolios (Jan 1997 - Dec 2013) 48
Table 5.7 Multivariate tests of normality 49
Table 5.8 Adjusted R-squared statistics of time-series regressions 52
Table 5.9 Adjusted R-squared statistic for cross-sectional regressions 55
Table 5.10 Mean absolute pricing error for cross-sectional regressions 58
Table 5.11 Factor risk premia per portfolio panel (Jan 1993 – Dec 2013) 60
Table 5.12 Factor risk premia per portfolio panel (Jan 1997 - Dec 2013) 63
Table 5.13 Asset pricing model tests 69

Figure 5.1 Scatter plots of expected return vs beta, coskewness and cokurtosis 66
Figure 5.2 P-Value plots where T = 120 71
Figure 5.3 P-Value plots where T = 252 72
Figure 5.4 Size-power plots where T = 120 73
Figure 5.5 Size-power plots where T = 252 74

1. Introduction

The Sharpe (1964) and Lintner (1965) Capital Asset Pricing Model (CAPM, henceforth) has served as one of the cornerstones of modern finance for half a century. The principle implication of the CAPM is a linear relationship between the expected return on assets and systematic risk. Systematic risk is defined in a linear pricing kernel by a single factor, the portfolio of aggregate wealth. The intuitive appeal of a linear relationship between risk and return cannot be denied. Unfortunately numerous studies have documented violations of this relationship in a large body of literature that has collectively become known as "anomalies". Anomalies are typically fundamental firm characteristics that account for a significant portion of cross-sectional variation in expected returns. The most well-known of these are the size effect, value effect and momentum effect[1].

Ross's (1976) arbitrage pricing theory (APT, henceforth) is the typical remedy for the failures of the CAPM. The addition of anomalies (or indeed other common risk factors) to the CAPM has become standard practice. Multifactor models have been successful in that they capture substantial portions of cross-sectional variation in expected returns. Fama and French (1993) introduce portfolios that proxy for size and value premia into the CAPM to create what became the most pervasive of these multifactor models. A common criticism of APT multifactor models relates to their foundations being based on statistical considerations making it difficult to ground the inclusion of the common risk factor in economic theory.

Assumptions of normality and linearity underpin the relationship between expected return and its covariance with common risk factors within APT type models. The validity of the assumption of normality has been in doubt from as early as 1963 when Mandelbrot illustrated excess kurtosis in U.S. equity returns. Multiple studies have documented the existence of skewness and kurtosis in asset return distributions (Fama, 1965; Rubenstein, 1973; and Kraus and Litzenberger, 1976; amongst others). The evidence of non-normality led to the introduction of higher moments into the CAPM, initially by Kraus and Litzenberger (1976) and more recently Harvey and Siddique (2000) and Dittmar (2002).

Harvey and Siddique (2000) and Dittmar (2002) extend the linear pricing kernel of the CAPM by the approximation of the first three polynomials of a Taylor series expansion to present non-linear APT type models. The quadratic and cubic CAPM models therefore include square and cubic terms of the portfolio of aggregate wealth, respectively. The existence of higher co-moments indicates that the investor consumption decision extends beyond the mean-variance efficiency framework. Expected return is then a function of an asset's covariance, coskewness and cokurtosis with the portfolio of aggregate wealth. The merging of the CAPM and non-linear APT type models is justified in utility theory under Kimball's (1990) concept of prudence. Harvey and Siddique (2000) introduce coskewness

[1] A comprehensive description of violations to the CAPM can be found in Fama and French (1992) or Campbell, Lo, and MacKinlay (1996).

under investor non-increasing absolute risk aversion while Dittmar (2002) introduces cokurtosis under investor aversion to extreme outcomes.

Skewness is generally accepted as priced in U.S. equity markets and the quadratic CAPM has been shown to capture additional variation in the cross-section of expected returns (Harvey and Siddique, 2000; Harvey, Liechty, Liechty and Muller, 2010). Skewness is for the most part shown to be negatively related to expected return returns (Harvey and Siddique, 2000; Harvey, Liechty, Liechty and Muller, 2010). There is less agreement on whether kurtosis is priced within US equity markets. Dittmar (2002) illustrates that preference restricted nonlinear pricing kernels are both admissible for the cross-section of returns and are able to significantly improve upon linear single- and multifactor kernels. This implies that higher moment CAPMs are superior in pricing assets when compared to traditional APT multifactor models such as the Fama and French (1993) three factor model (FF3F henceforth). Blau, Masud and Whitby. (2013) on the other hand find that when controlling for traditional risk factors, such as size, the excess return associated with excess kurtosis all but disappears.

The feasibility of the higher moment CAPMs have been well-documented for US equity markets but to the best of our knowledge no study has explored the relationship between expected return and higher moments in the South African equity market. The primary objective of this study is to determine the impact of relaxing the assumption of normality that underpins mean-variance efficient asset pricing. In particular, the study tests the feasibility of the quadratic and cubic CAPM models on Johannesburg stock exchange (JSE, henceforth) listed share return data. Higher moment models are compared to the linear CAPM and the Fama and French (1993) three factor model to determine which fit JSE listed equity data best.

A further objective of the study is to determine the impact on asset pricing model tests of the existence of higher moments within the sample data. Specifically, the study evaluates the appropriateness of test statistics used when faced with non-normal data within a South African context. For this the study introduces skewness and kurtosis into a Monte Carlo experiment designed to assess size and power characteristics of Ordinary Least Squares, Generalised Least Squares and Generalised Method of Moments estimation based test statistics.

The sample data consists of 53 portfolios weighted by market capitalisation (value weighted) and 20 equally weighted portfolios. JSE listed share returns are sorted into portfolios for two sample periods: (a) January 1993 to December 2013; and (b) January 1997 to December 2013. Sample portfolios are formed using either independent decile sorts or two-way quintile sorts on market equity (size) and book to market equity (value). The result is 10 portfolios formed on size, 10 portfolios formed on value and 25 portfolios formed on size and value. Finally industry portfolios are added to the sample for the January 1997 to December 2013 period. A longer sample period is preferred but we utilise the January 1997 and December 2013 period for two reasons: (1) a 60 month time- window was selected to estimate conditional higher moment models; and (2) return data for industry portfolios are only available starting January 1997.

The study makes multiple contributions to existing asset pricing literature. First, a distributional analysis of both value and equally-weighted portfolios is conducted. It is shown that skewness is

significantly different from that of a normal distribution for the majority of sample portfolios while kurtosis is significantly different from that of a normal distribution in all cases. Various test statistics are employed to establish non-normality on a univariate and multivariate basis for almost all of the sample portfolios. Collectively the results of the distribution analysis strongly suggests that JSE listed share returns are not normally distributed.

In the second contribution, the study employs a cross-sectional regression approach to estimate factor risk premia and test higher moment Capital Asset Pricing Models. Unconditional coskewness and unconditional cokurtosis are found to be priced within the size sorted and value sorted portfolio sets over the period January 1993 to December 2013. Conditional coskewness and conditional cokurtosis are found to be priced for only the size sorted portfolios over the period January 1997 to December 2013. This shows that the relationship between risk and return is considerably more complex than the conventional two dimensional perception. The signs on conditional coskewness and cokurtosis factors do not support the hypothesis that higher moments enter asset pricing based on investor prudence as Harvey and Siddique (2000) and Dittmar (2002) suggest. The result can be likened to Van Rensburg and Robertson's (2003) finding that beta is inversely related to return. The signs on the coskewness and cokurtosis factors show that investors are able to earn higher returns by investing in shares that are relatively less risky as defined by utility theory, violating the central premise of both the quadratic and the cubic CAPM.

The inverse relationship between risk and return is confirmed in the analysis of portfolios formed on coskewness beta as well. We document the spread of the positive coskewness portfolio (S^+) over the negative coskewness portfolio (S^-) for the period January 1997 to December 2013 at -0.97 percent per month on value weighted portfolios. This result is significant in that it casts doubt on the inclusion of coskewness in the quadratic CAPM based on investor non-increasing absolute risk aversion. Furthermore this result illustrates that a dynamic trading strategy based coskewness can lead to abnormal profits. An arbitrage portfolio of a long position in S^+ and a short position in S^- would have earned an investor an annualised return of 12.28 percent between January 1997 and December 2013.

The results of the asset pricing model tests are less clear. The pricing errors for higher moment Capital Asset Pricing Models are shown to be statistically significant for size sorted portfolios while pricing errors on the value sorted, size-value sorted and industry portfolios are found to be statistically insignificant. The study thus highlights the dependence of results on the firm characteristic used in portfolio formation. Also, the relative superiority of the Fama and French (1993) three factor model over the January 1993 to December 2013 period casts doubt on Dittmar's (2002) assertion regarding superiority of the preference restricted non-linear pricing kernels over multifactor pricing kernels. An holistic assessment of the results of the asset pricing test statistics lead to the conclusion that none of the models tested are likely the true model. The true model should explain variation in returns regardless of the data generating process.

Additionally a periodic analysis of the Fama and French (1993) size and value factors provides two interesting results. Firstly we find that the size effect is concentrated in the January 1993 to December 1996 sub-period, being three to four times larger than the overall sample. Secondly we show a reversal

of the size and value effect between January 2008 and December 2013. We hypothesise that the early concentration of the size and value premia, combined with the recent reversal are the causes of the poor performance of the FF3F model over the January 1997 to December 2013 sample period. These findings are confirmed in the cross-sectional regressions as the Fama and French (1993) size and value factors are priced when estimated over January 1993 to December 2013 but lose their significance when estimated over the period January 1997 to December 2013.

Finally, the study addresses the impact of higher moments on the testing methodology. The results of the Monte Carlo experiment show that an Ordinary Least Square Wald test statistic has the most desirable size characteristic while a Generalised Least Squares J-test statistic has the most desirable power characteristic when dealing with non-normal data. Cochrane (2005) suggests using Hansen's (1982) test of overidentifying restrictions when testing asset pricing models estimated by the Generalised Method of Moments. The results of this study suggest that the test of overidentifying restrictions has undesirable size and power characteristics.

This study is structured as follows: Section Two reviews pertinent previous literature; Section Three describes the sample data; Section Four outlines the methodology employed; Section Five summarises empirical results; and Section Six concludes.

2. Literature Review

This section commences with a brief outline of linear factor asset pricing model notation and concepts. This is followed by a summarised discussion of the literature pertaining to the most prominent asset pricing theories in economic literature.

2.1 Asset Pricing Models

The most popular asset pricing models in finance are undoubtedly linear factor models. Linear factor models are based on the following condition:

$$E\left[\left(1 + R^i_{t+1}\right)m_{t+1}\right] = 1 \qquad (2.1)$$

Where $\left(1 + R^i_{t+1}\right)$ is the total return on asset i at time $t + 1$, and m_{t+1} is the marginal rate of substitution. The marginal rate of substitution refers to the willingness of an investor to exchange consumption at time $t + 1$ for consumption at time t. Investors will only hold an asset should Equation 2.1 be satisfied.

The marginal rate of substitution is often referred to as a stochastic discount factor or pricing kernel, it is not directly observable. Asset pricing models most often develop proxies for the marginal rate of substitution that take the form of linear factors. Different asset pricing models differ primarily in the manner in which these proxies are defined. Examples of such proxies are firm characteristics, macroeconomic variables, returns on financial assets and so forth. The following sections explore some of these asset pricing models and their corresponding proxies. The pricing kernel in factor models can be expressed as:

$$m_{t+1} = a + b' f_{t+1} \qquad (2.2)$$

Where a and b are parameters and f is a vector of factors. Time-series regressions are often utilised to estimate factor loadings such that:

$$R^i_t = a^i + \beta_i f_t + \varepsilon^i_t \qquad t = 1, \dots, T \qquad (2.3)$$

Where R^i_t is the return on asset or portfolio i at time t; a^i is the time-series regression intercept; β_i is a $1 \times K$ vector of regression coefficients, K being the number of factors; and ε^i is the time-series regression residual. Factor models can be expressed in expected return-beta form:

$$E\left(R^i\right) = \beta^i E(f) + \alpha^i \qquad i = 1, \dots, n \qquad (2.4)$$

Where the β^i are the time-series regression coefficients and α^i is the pricing error of the asset pricing model. In most cases α is assumed to be equal to zero.

2.2 Capital Asset Pricing Model

The pioneering work by Harry Markowitz (1952, 1959) in portfolio theory has been described as the roots of modern asset pricing theory. Markowitz (1952) hypothesised that investors view returns as desirable and variance as undesirable. This basic idea serves as the foundation for his concept of mean-variance efficiency. Mean-variance efficiency involves the manner in which risk-averse investors

construct portfolios to optimise or maximise expected return based on a given level of systematic risk. Portfolios are considered mean-variance efficient when a higher return is not possible at a given level or risk. Markowitz christened the locus of efficient portfolios available to investors the *efficient frontier.*

Sharpe (1964), Lintner (1965), Mossin (1966) and Treynor (1963) extended Markowitz's work to an equilibrium asset pricing theory. Central to extended model is the addition of the risk-free rate first introduced by Tobin (1958). The risk free rate refers to an asset that offers a constant return in all states of the world to the efficient frontier. The introduction of the risk-free rate might seem trivial but it allows investors to discard all but one portfolio of risky assets. Sharpe (1964) illustrates that all rational investors would seek to hold differing combinations of the portfolio of risky assets and the risk free rate depending on their risk preferences. Graphically this can be represented by a tangent line from the intercept point on the efficient frontier to the point where the expected return equals the risk-free rate of return. This is commonly referred to as the capital markets line. Assuming efficient markets, clearing prices require this tangent portfolio to be a composite of all risky assets in the economy. This composite portfolio of all risky assets is generally referred to as the market portfolio.

Portfolio theory illustrates that idiosyncratic risk can be diversified away by the addition of securities to a portfolio, leaving portfolios at risk only to market factors (Markowitz, 1952). The CAPM builds on this idea suggesting that the risk of any security is not a function of its standard deviation but rather its covariance with the market portfolio. The CAPM's pricing kernel is:

$$m_{t+1} = a + bR^m_{t+1} \tag{2.5}$$

Where R^m is the return on the market portfolio. The CAPM can be expressed in expected-return beta form as:

$$E(R^i) = \beta_i E(R^m) \tag{2.6}$$

Assuming the existence of a risk free asset or Black's (1972) zero beta portfolio the CAPM becomes:

$$E\left(R^i\right) = R^f + \beta_i[E(R^m) - R^f] \tag{2.7}$$

Where R^f is the return on the risk free asset or zero beta portfolio. The CAPM therefore explicitly states that expected return is a function of β and the excess return on the market portfolio over the risk-free rate. This excess return is usually denoted as the market risk premium. β is calculated by dividing the covariance of asset i's returns with the return of the market portfolio by the variance of the market portfolio:

$$\beta_i = \frac{Cov(R^{ei}, R^{em})}{Var(R^{em})} \tag{2.8}$$

Where R^e refers to an asset or portfolio's return over the risk free rate. β is a measure of the risk arising from exposure to general market movements as opposed to idiosyncratic factors. Equation 2.8 serves as a formal definition of systematic risk. The CAPM thus posits a linear relationship between risk and return, i.e. investors are only able to earn higher returns by investing in securities with higher betas.

Assumptions

Central to the CAPM is a set of assumptions that are frequently described as unrealistic, they are: (1) All investors are rational risk averse utility maximisers; (2) All investors have homogeneous expectations; (3) Investors are able to lend and borrow unlimited amounts at the risk free rate; (4) No transaction or taxation costs exist; (5) Investors are able to take limitless long and short positions across assets; (6) Investors are price takers; (7) Investors have a single period investment horizon; and finally (8) Asset returns are jointly normally distributed.

The assumptions are often likened to a physicists vacuum. The fact that they do not always hold in practice, does not necessarily imply that implications of the CAPM are null and void. Many extensions of the CAPM have addressed the practicality of its underlying assumptions. Black (1972) presents the zero beta CAPM that that relaxes assumption of the existence of riskless borrowing and lending opportunities. Mayers (1972) develops a version of the CAPM that extends the market portfolio proxy to include human capital. Merton (1973) derives an intertemporal CAPM that extends the mean-variance efficiency to long-term wealth effects. Williams (1977) derives a complex version of the CAPM that accommodates heterogeneous beliefs. His risk return-relationship is based on individual wealth, risk aversion, and prior beliefs held by each of the investors in the market. Extensions of the CAPM are well-documented in finance and economics literature. This study primarily focusses on the final assumption of multivariate normal return distributions as it lies in stark contrast to the assumptions underpinning the asset pricing models explored later.

The Conditional CAPM

The Conditional CAPM is an important extension for the discussion later and as such a brief overview of the model is given. Thus far the focus has been on models where the factor exposures remain constant over time. In reality linear factor models include parameters that vary as function of both time and conditioning information (Chan and Chen, 1988). Simply put this means that information accumulates in financial markets and market participants make investment decisions contingent on this information available at the time. The first order condition for an investor holding a risky asset then becomes:

$$E\left[\left\langle\left(1+R_{t+1}^{i}\right)m_{t+1}\middle|\Omega_{t}\right\rangle\right]=1 \tag{2.9}$$

In this version total return and the marginal rate of substitution are conditional on information set Ω, available to investors at time t. The primary implication of conditional information for capital asset pricing is time-varying market betas. Equation 2.7 becomes:

$$E\left(R_{t}^{i}\right)-R^{f}=\beta_{t}^{i}\left[E(R_{t}^{M})-R^{f}\right] \tag{2.10}$$

Apart from the subscript on β, meant to emphasize time-variability, the model is very similar to Equation 2.7. The central premise is that conditional alpha is always zero with pricing errors resulting from the time-variation in betas (Jensen, 1968; Jagannathan and Wang, 1996). Any asset pricing model can be, and usually is, extended to conditional form.

Early Tests of the CAPM

The CAPM is arguably the most scrutinised model in finance. Initial tests of the CAPM provided support for its central hypothesis. In a seminal study Fama and Macbeth (1973) develop a methodology to test the empirical validity of the CAPM. They find moderate support for the CAPM and that their market proxy the NYSE index is consistent with efficiency. The Fama-Macbeth (FM henceforth) technique consists of two sets of regressions. First rolling window time-series regressions are used to calculate a series of market betas. The cross-sections of one period ahead portfolio/stock returns are then regressed on these time-series betas to estimate market risk premia. Finally, risk premia are averaged over cross-sections to determine final parameter values. The FM technique is still widely used today and will be described in depth later in this study.

Other early tests did not support the risk-return relationship proposed by the CAPM. Friend and Blume (1970) found that individual asset betas are imprecise. Black, Jensen and Scholes (1972) utilised a two-stage testing methodology. They estimated market betas using time series regressions of the monthly returns of stocks listed on the NYSE, over the 1926-1930 period, on an equally-weighted portfolio composed of all stocks on the NYSE. Black, Jensen and Scholes (1972) show that the CAPM does not hold in practice for the specified sample period.

Finally, Roll's (1977) criticism of asset pricing tests should also be highlighted. He argues that testing any two parameter asset pricing model is difficult and probably infeasible for the foreseeable future. According to Roll the only testable hypothesis within an asset pricing model is that the market portfolio is mean-variance efficient. None of the other implications (i.e. a linear relationship between expected return and systematic risk) can be tested independently since they depend on the market portfolio being mean-variance efficient. The crux of his argument is that any valid test should use the true market portfolio composition, i.e. every possible asset needs to be included. Roll's criticism is overly harsh and suggests that you cannot make any positive statements about asset pricing models unless the true market portfolio is known. Roll's (1977) criticism did not prevent other researchers from finding more inventive tests of the CAPM. These tests involve the use of trading strategies to identify contradictions to the linear risk-return relationship suggested by the CAPM. Contradictions which collectively became known as anomalies.

Anomalies to the CAPM

The CAPM and Fama's (1970) Efficient Market Hypothesis (henceforth EMH) are often referred to as joint hypotheses. Conclusions drawn from tests of these complimentary hypotheses are difficult to separate empirically, a phenomenon often referred as the joint hypothesis problem. In short the EMH suggests that the market quickly and correctly incorporates all relevant information resulting in fairly priced securities. The idea is that numerous investors trade based on the information available to them and in doing so they incorporate this information into the market. As more investors trade the profit or arbitrage opportunities that motivate the trade rapidly disappear. Efficiency can then be attributed to the investors' profit motive and subsequent competition. Frictionless markets will therefore assimilate information instantaneously.

Accordingly the only manner in which investors can increase their profit within an efficient market is by assuming more risk, regardless of the information they hold. Yet many studies have shown the existence of anomalies to the CAPM's systematic risk-return paradigm. These anomalies suggest the ability to obtain consistent abnormal profits from various dynamic trading strategies. The most salient of these anomalies are the January effect, the size effect, the value effect and the momentum effect.

The January effect was first identified by Rozeff and Kinney (1976). It suggests that the risk adjusted returns in January are abnormally higher than in any other month. Rozeff and Kinney (1976) examined the existence of seasonal patterns in an equally weighted index of the NYSE from 1904 to 1974. They find that the mean of January returns is significantly higher than in any other month, confirming the existence of January effect. This result has been confirmed by Keim (1983), Reinganum (1981) and Roll (1983) amongst others. Interestingly these studies found that the January effect is concentrated in firms with low market capitalisation. In fact, the January effect seems to be absorbed by the size effect.

The size effect was first demonstrated by Banz (1981) in his paper "The relationship between return and market value of common stocks". The size effect proposes that portfolios composed of low market capitalization shares outperform those composed of high capitalization shares on a risk adjusted basis. Banz (1981) investigated the relationship between the total market value of common stocks listed on the NYSE and their returns for the period 1936 - 1975. Banz (1981) documents that stocks in the quintile portfolio with the smallest market capitalisation earn a risk-adjusted return that is 0.40% higher per month than in other quintiles. His results have been corroborated by, although at different magnitudes, by Reinganum (1981), Brown, Kleidon, and Marsh, (1983) Keim (1983) and Fama and French (1992).

The value premium or effect was initially documented by Basu (1977). It posits a relationship between fundamental ratios (a firm characteristic to price) and return. Basu (1977) found that for the period April 1957 to March 1971, low price to earnings (P/E) ratio securities on average earned higher absolute risk adjusted returns compared to high P/E ratio securities. The value effect is not limited to shares with low P/E ratios. Fama and French (1992) show that the book equity (BE) to market equity (ME) ratio, combined with the size effect, subsumes the premium attributable to other value indicators. They find that portfolios consisting of high BE / ME outperform portfolios consisting of low BE / ME shares. The value effect has further been corroborated by Rosenberg, Reid and Lanstein (1985), Chan, Hamao and Lakonishok (1991) and Lakonishok, Shleifer and Vishny (1994).

The momentum effect was first identified by Jegadeesh and Titman (1993). Momentum suggests that securities' prices are more likely to keep moving in the same direction than to change direction. A momentum-based trading strategy therefore would buy past 'winners' and sell past 'losers'. Jegadeesh and Titman (1993) find that trading strategies that buy past winners and sell past losers realize significant abnormal returns over the 1965 to 1989 period. The momentum effect has been corroborated by Bhoraj and Swaminathan (2006), Lewellen (2002) and Nijman, Swinkels and Verbeek (2004) amongst others.

The constant factor exposures in the unconditional CAPM is often said to be the cause of these anomalies. There are several studies that show that time-variation in factor loadings help explain the

size, value and momentum effects. Zhang (2005) presents a model where value stocks are riskiest in bad times and the relationship between parameter and the risk premium leads to an unconditional value premium while conditional CAPM alphas are zero. Results similar to these are not new to the literature. For a duration of time the conditional CAPM was heralded as the saviour of the unconditional CAPM. Primarily these studies show that the size and value effects are not robust when accounting for business cycles (Jagannathan and Wang, 1996; Lettau and Ludvigson, 2001; Petkova and Zhang, 2005; Avramov and Chordia, 2006; Wang, 2002; Ang and Chen, 2007; Lustig and Van Nieuwerburgh, 2003).

Lewellen and Nagel (2006) argue that significant departures from the unconditional CAPM would require implausibly large time-variation in betas and expected returns. The conditional CAPM as such would not explain asset-pricing anomalies like book-to-market and momentum. They estimate conditional alphas and betas from short window regressions and show that the conditional CAPM performs nearly as poorly as the unconditional CAPM.

South African Anomalies

Apart from the January effect, the existence of these anomalies have been well-documented within the South African context. Van Rensburg and Robertson (2003), in a study of Johannesburg Stock Exchange (JSE) listed firm data from July 1990 to June 2000, show that small size and low price to earnings ratio stock portfolios have higher comparative returns while having lower comparative betas. This suggests a negative relationship between risk and return i.e. an investor can achieve a higher return by selecting “less risky” stock. Their results directly contradict the central premise of the CAPM, a positive linear relation beta systematic risk and expected return. This does not bode well for the CAPM within the South African context.

In a more comprehensive study Basiewicz and Auret (2009) employ FM regressions combined with annual univariate and multivariate portfolio sorts on size and value indicators to both equally weighted and value weighted portfolios for JSE listed firms from July 1992 to July 2005. Echoing Fama and French (1992) they confirm that BE/ME is a superior value indicator compared to the P/E ratio utilised by Van Rensburg and Robertson (2003). Basiewicz and Auret (2009) additionally confirm the existence of the size and value effects for both equally weighted and value weighted portfolios. For univariate portfolio sorts the size effect on the equally weighted portfolios is documented at 1.1 percent per month and 0.87 percent per month on the value weighted portfolios. While the value effect on the equally weighted portfolios is document at 1.56 percent per month and 1.5 percent per month on the value weighted portfolios. Finally Basiewicz and Auret (2009) confirm the existence of an inverse relationship between beta and expected return. Hoffman (2012) confirms the momentum effect on the JSE listed firm data from 1985 to 2010 and documents its magnitude at between 1.4 and 2.5 percent per month.

There is some disagreement in the literature as to whether the size and value effect actually exist on the JSE. Auret and Cline (2011) extend a study done by Robins, Sandler and Durand (1999) in which they investigated whether or not the inter-relationships between the value, size and January effects can be detected on the JSE. They use two sample periods, January 1988 to December 1995 and January

1996 to December 2006. Like Robins *et al*. (1999) Auret and Cline (2011) find no significant size or value effects in either of the periods.

By construction the methodology of this study allows for the evaluation of the size and value effects over a long sample period by South African standards. Even though not the main of the paper, this study provides clarity on the existence of the size and value effects on the JSE and their relationships with higher moments.

2.3 Arbitrage Pricing Theory

Ross (1976) developed arbitrage pricing theory (APT, henceforth) which suggest that asset pricing models should be based on statistical considerations and be justifiable on grounds. The fundamental idea underpinning APT is based on portfolio theory. Since idiosyncratic risk can be diversified away one should be able to relate a security's expected return to its covariance with common risk factors. The factors are typically related to asset returns by linear regression. Arbitrage pricing theory suggests that an asset's return can be dependent on multiple factors e.g. macroeconomic variables or market indices. The pricing kernel in Equation 2.2 can thus be ascribed to APT models:

$$m_{t+1} = a + b' f_{t+1} \tag{2.2}$$

While Equation 2.4 is the APT expressed in expected return beta form:

$$E\left(R^i\right) = \beta^i E(f) + \alpha^i \tag{2.4}$$

APT models start with the statistical characterisation of outcomes in order to derive a model's expected return. A well-known example is the macroeconomic variable APT of Chen, Roll and Ross (1986). They model equity returns as functions of macroeconomic variables and non-equity asset returns to determine if their risks that are rewarded by the stock market. Interestingly they find term spread, unexpected inflation, industrial production and spread between high and low grade bonds are significantly priced sources of risk while neither the market portfolio, consumption nor oil prices are priced separately on the stock market. The Fama and French (1993) three factor model is likely the most pervasive APT model in finance and is still commonly used today.

2.4 Fama French Three Factor Model

In their seminal work Fama and French (1992) investigate the joint roles of market β, size, Earning/Price (E/P), leverage, and BE/ME in the cross-section of average stock returns. They employ FM regressions on all non-financial firms in the intersection of the NYSE, AMEX and NASDAQ return files from the Centre for Research in Security Prices (henceforth, CRSP) as well as the Compustat database. Their results completely contradict the CAPM. The authors find that, whether used alone or in combination with other variables, β has little information about the cross-section of average returns. Furthermore they find that size and book-to-market equity seem to absorb the apparent roles of leverage and E/P in average returns. Based on these findings Fama and French (1993) present the three factor model based on the following pricing kernel:

$$m_{t+1} = a + bR^m_{t+1} + cSMB_{t+1} + dHML_{t+1} \tag{2.11}$$

Fama and French (1993) construct portfolios based on independent two-way sorts of two size portfolios on three value portfolios. SMB (size factor) is the series of average returns on the three small portfolios minus the average return on the three big portfolios; and HML (value factor) is the average return on the two value portfolios minus the average return on the two growth portfolios. Theoretically the FF3F model implies that there are sources of priced risk beyond the market factor as suggested by the CAPM. These sources of common risk are captured in the proxy portfolios SMB and HML.

There are multiple theoretical explanations as to why the FF3F model factors explain variation in average returns unrelated to market beta. Fama and French (1996) explain that high BE / ME stocks or value stocks have suffered a succession of bad news driving the price down. Firms are in or near financial distress. A trading strategy that entails purchasing firms on the verge of bankruptcy has experienced higher returns as firms come out of bankruptcy more often than not. Firm value thus proxies for financial distress. Importantly Cochrane (2005) notes that one cannot count an individual firm event, as all idiosyncratic risk can be diversified away, only aggregate market events. For example financial crises result in a flights to quality which in turn result in financial distress stocks performing very poorly. The fact that the HML portfolio does not covary strongly with other measures of financial distress provides limited support for Fama and French's (1996) hypothesis.

Another commonly expressed explanation is that the size and value premia proxy for a liquidity premium. Investors receive or demand a liquidity premium when the security they hold cannot be easily converted into cash at its fair value. Very small firms and firms that are close to financial distress (value firms) are not likely to be traded as often as larger or financially sound firms. There is thus an increased risk of not being able sell small or value firms during a market downturn, hence investors demand a premium to hold such firms. Liu (2006) finds support for this hypothesis when he constructs a Liquidity Augmented CAPM and finds that size and value premia are subsumed by his new liquidity measure. Differential information, seasonality and transaction costs[1] have also been put forward as causes of the size or value premium.

The FF3F model is most often expressed as:

$$R_t^i - R_t^f = a^i + \beta_i^m R_t^{em} + \beta_i^s SMB_t + \beta_i^v HML_t + \varepsilon_t^i \qquad (2.12)$$

Equation 2.1 represents a time-series regression where excess returns for asset i are regressed on market, size and value factors at time t. The β's are regression coefficients for the market, size and value factors.

[1] Refer to Fama and French (2006, 2008) and De Moor and Sercu (2013) for a more comprehensive discussion on size and value premium explanations.

Table 2.1 Extract from Fama and French (1993, p. 20 & 25) results

The table presents the R^2 of the CAPM and FF3F model time-series regressions completed in Fama and French (1993). They construct portfolios at the breakpoints between independent quintile sorts on ME and BE/ME of CRSP shares for period July 1963 to December 1991.

CAPM

	BE/ME quintiles				
Size quintiles: R^2	Low	2	3	4	5
Small	0.67	0.7	0.68	0.65	0.61
2	0.79	0.79	0.76	0.76	0.71
3	0.84	0.81	0.8	0.79	0.74
4	0.89	0.9	0.87	0.8	0.76
Big	0.89	0.91	0.84	0.79	0.69

FF3F

	BE/ME quintiles				
Size quintiles: R^2	Low	2	3	4	5
Small	0.94	0.96	0.97	0.97	0.96
2	0.95	0.96	0.95	0.95	0.96
3	0.95	0.94	0.93	0.93	0.93
4	0.94	0.93	0.91	0.89	0.89
Big	0.94	0.92	0.88	0.9	0.83

Fama and French (1993) use the time-series regression approach of Black, Jensen, and Scholes (1972). They construct portfolios at the breakpoints between independent quintile sorts on ME and BE/ME for the period July 1963 to December 1991. The 25 portfolio returns are then regressed on market, size and value factors. Fama and French (1993) obtain R-squared values that are greater than 0.9 in 21 of the 25 regressions. For comparison Fama and French (1993) run regressions with only the market factor, i.e. the CAPM, for which they find only two R-squared values above 0.9. An excerpt of their results is presented in Table 2.1. Their results indicate that approximately 90 percent of the variation in stock returns can be explained by the addition of size and value factors to the CAPM.

Table 2.2 Extract from Basiewicz and Auret (2010, p. 19 & 22) results

The table presents the R^2 of the CAPM and FF3F model regressions completed in Basiewicz and Auret (2010). They construct portfolios at the breakpoints between independent sorts of JSE listed shares into four ME and three BE/ME portfolios for period July 1992 to July 2005.

CAPM

		BE/ME quintiles		
Size quintiles	R^2	Low	2	High
	Small	0.522	0.497	0.416
	2	0.57	0.575	0.481
	3	0.88	0.867	0.577
	Big	0.358	0.452	0.473

FF3F

		BE/ME quintiles		
Size quintiles	R^2	Low	2	High
	Small	0.729	0.708	0.546
	2	0.657	0.718	0.524
	3	0.89	0.876	0.583
	Big	0.395	0.469	0.595

Basiewicz and Auret (2010) assess the feasibility of the FF3F model on the JSE. They construct portfolios at the breakpoints between independent sorts of JSE listed shares into four ME and three BE/ME portfolios for the period July 1992 to July 2005. The FF3F model substantially improves on the CAPM in most portfolios but the R-squared values are much lower compared to those in Table 2.1. Not a single value is in excess of 0.9. The FF3F model therefore appears less able to explain time-series variation in South African equity market returns.

The FF3F model can be expressed in expected return beta form as:

$$E\left(R^i\right) - R^f = \beta_i^m \lambda^m + \beta_i^s \lambda^s + \beta_i^v \lambda^v \qquad (2.13)$$

The β's are regression coefficients estimated using Equation 2.13 while the λ's are the premia associated with risk factors for which we use the market, size and value factors as proxies. λ's can be calculated by regressing the cross-section of expected returns on parameters identified using Equation 2.13. The R-squared values of cross-sectional regressions are typically lower than that the R-squared statistics resulting from time-series regressions. The FF3F model typically results in R-squared values in excess of 0.70 when expected returns are regressed on market, size and value parameters (Ferguson and Shockley, 2003; Hahn and Lee, 2006; Petkova, 2006).

A well-known extension of the FF3F model, is the Carhart (1997) four factor model which incorporates the Jegadeesh and Titman (1993) momentum effect. The proxy for the momentum factor is constructed, in a similar manner as that size and value factor proxies, using the difference between

the returns on the winners' portfolio and the returns on the losers' portfolio for a given universe of equities. The Carhart (1997) model can be expressed mathematically as:

$$R_t^i - R_t^f = a^i + \beta_i^m R_t^{em} + \beta_i^s SMB_t + \beta_i^v HML_t + \beta_i^{momentum} WML_t + \varepsilon_t^i \quad (2.14)$$

Where WML_t is the proxy for the momentum factor at time t. Fama and French (2012) form 25 size and BE/ME sorted portfolios as well as 25 size and momentum portfolios for the period November 1989 to March 2011 at a regional level for North America, Europe, Japan, and Asia Pacific. Fama and French (2012) find that four factor asset pricing models are "rather successful" in capturing expected return at a regional level for the size-BE/ME portfolios but are less successful for the size-momentum sorted portfolios. Furthermore they find that the four-factor model performs as well or better than the three-factor model or the CAPM at regional level.

2.5 Higher Moment Capital Asset Pricing Models

The models discussed thus far have all been linear factor models that assume asset returns follow a joint normal distribution. The models by definition assume that the higher moments conform to that of a normal distribution. Skewness is the third moment of a distribution. It characterises the degree of asymmetry of the distribution around its mean. The third moment of a distribution is calculated by:

$$Skew = \frac{1}{T}\sum_{t=1}^{T} \frac{\left(R_t^i - E[R^i]\right)^3}{\left(\sigma^i\right)^3} \quad (2.15)$$

A normal distribution is symmetric i.e. not skewed ($Skewness = 0$). Positive skewness exists when the distribution is asymmetric with a tail favouring positive values while negative skewness is said to exist when the distribution is skewed towards negative values.

Kurtosis is the fourth moment of a distribution. It describes the relative peakedness or flatness of a distribution directly compared to that of a normal distribution. The forth moment of a distribution is calculated by:

$$Kurt = \frac{1}{T}\sum_{t=1}^{T} \frac{\left(R_t^i - E[R^i]\right)^4}{\left(\sigma^i\right)^4} \quad (2.16)$$

The kurtosis value of a normal distribution is three. Distributions are described as leptokurtic when kurtosis is in excess of three and thus more peaked than a normal distribution. Distributions are described as platykurtic when kurtosis is below three which indicates a comparatively flat distribution.

Mandelbrot (1963) showed that stock returns are too peaked to be considered normally distributed, a finding Fama (1965) confirmed. Harvey and Siddique (2000) construct the 25 Fama and French (1992) size and BE/ME sorted portfolios using NYSE/AMEX shares for the period July 1963 to December 1993 and find that skewness is significant at the 10 percent level 17 out of 25 times. Vorkink (2003) constructs decile portfolios using independent sorts on ME and momentum using CRSP monthly returns (NYSE, NASDAQ, and AMEX) for July 1963 to December 1995. The author finds that 13 of the 20 portfolios are significantly skew and that all portfolios contain kurtosis that are significantly

different from a normal distribution. Equity returns within developed markets are generally accepted to be non-normal.

Hwang and Satchell (1999) evaluate data from January 1985 to January 1997 for 17 emerging market countries, excluding South Africa, and find evidence for significant skewness and kurtosis. Within the South African context, Mangani (2007) analyses JSE All Share Index returns from December 1983 to April 2002 and 42 individual stock returns from February 1973 to April 2002. The author found skewness statistically insignificant in only six of the forty two shares while sample kurtosis was significantly greater than three in virtually all the cases, and generally very large.

The existence of excess kurtosis in asset return distributions would indicate that extremely large price movements occur more often than predicted by a normal distribution. The frequency of market crashes and bubbles over the past century (i.e. negative tail events) certainly provide a strong case for the existence of excess kurtosis within asset returns. Xiong and Idzorek (2011) actually show that extremely large price movements occur approximately 10 times more often than predicted by a normal distribution for 14 different asset classes. The existence of positive skewness in the return distribution of a specific asset would indicate that the asset was more prone to large positive price movements as oppose to large negative price movements. Conversely, the existence of negative skewness in the return distribution of a specific asset would indicate that the asset was more prone to large negative price movements.

Skewness in Asset Pricing

The effect of skewness is not new to asset pricing literature. A higher moment CAPM was originally proposed by Rubenstein (1973) after noting that US security returns are skewed or leptokurtic. Building on Rubenstein's work, Kraus and Litzenberger (1976) introduce the third moment of return distributions into the unconditional linear CAPM. They develop the concept of systematic skewness or coskewness in what has become a seminal work. The fundamental justification of the quadratic CAPM is based on an investor's non-increasing absolute risk aversion.

Harvey and Siddique (2000) relate non-increasing absolute risk aversion for a risk-averse utility-maximizing investor to Kimball's (1990) concept of prudence. Prudence refers to the desire to avoid disappointment and is associated with the precautionary savings motive. Non-increasing absolute risk aversion suggests that when investors hold a portfolio of risky assets, increases in the total skewness of the portfolio are preferred. The addition of an asset with negative coskewness to this portfolio would result in a more negatively skewed portfolio i.e. reduces the total skewness of the portfolio. Investors would therefore require a premium to hold assets with negative coskewness all other things being equal. One thus expects a negative relationship between systematic skewness and average returns when evaluating cross-sections of assets. Furthermore the premium for skewness risk over the risk-free rate should also be negative (Harvey and Siddique, 2000).

This concept is best illustrated by considering negatively and positively skewed return distributions and investor preferences. Positively skewed assets have a higher probability of extreme positive events while negatively skewed assets have a higher probability of extreme negative events. Logically

investors should prefer assets that are positively skewed as opposed to negatively skewed. Harvey e*t al.* (2010) summarise an aversion towards negatively skewed returns as the basic intuition behind investors willingness to trade some of their average return for a decreased chance that they will experience a large reduction in their wealth. If an asset then decreases a portfolio's skewness it should provide a higher return to accommodate for the additional risk. Similarly, one can expect assets that increase a portfolio's skewness to have lower expected returns.

What does this mean for mean-variance efficiency and the efficient frontier? The inclusion of skewness in asset pricing models indicates that mean and variance cannot adequately characterise investor preferences as originally proposed by Markowitz (1952). As Harvey and Siddique (2000) graphically illustrate, at any level of variance, there is a negative trade-off between expected return and skewness. They develop a three-dimensional "efficient frontier" where there are multiple efficient portfolios. The capital market line is extended to the capital market plane and optimal portfolios are at tangency points of the investor's indifference surface. Harvey *et al.* (2010) find that a utility function approximated by a third order Taylor series expansion leads to more informatively selected portfolio weights as they incorporate the effects of skewness. The authors explain this concept using a two stock portfolio. The portfolio mean is identical to the linear combination of the stock means and the portfolio variance is less than the combination of the stock variances but there is no guarantee that the portfolio skewness will be larger or smaller than the linear combination of the stock skewness. This suggests that the mean-variance optimal portfolios will likely result in sub-optimal portfolios in the presence of skewness.

Mitton and Vorkink (2007) develop a one period asset pricing model that incorporates investors' heterogeneous preference for skewness. The central premise of the model is that heterogeneous preference for skewness allows investors in equilibrium to under diversify. Investors are willing to sacrifice some-mean variance efficiency within their portfolio for an increased possibility of a large positive return i.e. positive skewness. Barberis and Huang (2008) justify the inclusion of skewness into their asset pricing model under Tversky and Kahneman's (1992) cumulative prospect theory. The principle effect of the cumulative prospect theory weighting function is to overweigh the tails of the distribution when making investment decisions. Barberis and Huang (2008) believe that the overweighting of tails captures investor preference for a lottery-like assets, or asset with a positively skewed return distribution.

If, in the three moment CAPM, expected return is function of an asset's covariance and coskewness with the market portfolio. Then three moment CAPM's[1] marginal rate of substitution is:

$$m_{t+1} = a + bR_{t+1}^{m} + c(R_{t+1}^{m})^2 \qquad (2.17)$$

[1] For a detailed derivation of the three moment CAPM's marginal rate of substitution grounded in utility theory, please refer to Harvey and Siddique (2002).

Accordingly the Kraus and Litzenberger (1976) CAPM is:

$$R^i - R^f = \alpha^i + \beta_i^m\left(R^m - R^f\right) + \beta_i^{m2}\left(R^m - R^f\right)^2 + \varepsilon^i \quad (2.18)$$

Where β_i^{m2} is asset i's coskewness with the market portfolio. β_i^{m2} is calculated by:

$$\beta_i^{m2} = \frac{Cov\left[R^{ei}(Re^m)^2\right]}{E\left[\left(R^{em} - E(R^{em})\right)^3\right]} \quad (2.19)$$

The model can equivalently be expressed in expected return beta form:

$$E\left(R^i\right) - R^f = \beta_i^m\lambda^m + \beta_i^{m2}\lambda^{m2} \quad (2.20)$$

Where the λ's are factor risk premia. Harvey and Siddique (2000) extend the quadratic CAPM introducing conditional systematic skewness where a, b and c in Equation 2.18 are functions of the period t information set Ω. Mathematically their model is:

$$R_t^i - R_t^f = \alpha_t^i + \beta_i^m R_t^m + \beta_{i,t}^{sks}\left(R_t^m - R_t^f\right)^2 + \varepsilon_t^i \quad (2.21)$$

Conditional coskewness is then calculated as[1]:

$$\beta_{i,t}^{sks} = \frac{E\left(\varepsilon_{i,t+1}\varepsilon_{m,t+1}^2\right)}{\sqrt{E\left(\varepsilon_{i,t+1}^2\right)}E\left(\varepsilon_{m,t+1}^2\right)} \quad (2.22)$$

Where $\varepsilon_{i,t} = R_{t+1}^{ei} - \alpha_t^i + \beta_{i,t}^m R_t^{em}$. Harvey and Siddique (2000) construct 25 size-BE/ME sorted portfolios as well as 10 momentum sorted portfolios using NYSE and AMEX sharers for July 1993 to December 1993. They then use FM regressions to contrast the CAPM, conditional quadratic CAPM and the FF3F model. As previously illustrated the FF3F model almost always does better than the unconditional CAPM in explaining variation in expected return. Harvey and Siddique (2000) show that the addition of conditional coskewness makes the single-factor model strikingly more competitive. For the size-BE / ME sorted portfolios the R-squared are as follows: (1) the unconditional CAPM – 0.11; (2) the FF3F model – 0.72; and (3) the conditional quadratic CAPM – 0.68. The results found on the momentum portfolios are: (1) the unconditional CAPM – 0.04; (2) the FF3F model – 0.89; and (3) the conditional quadratic CAPM – 0.61. Finally they add a systematic skewness factor to the FF3F model and find that the R-squared increases to 0.83 and 0.96 for the size and momentum portfolios respectively. Harvey and Siddique's (2000) results clearly illustrate that conditional skewness helps explain the cross-sectional variation in expected returns of size-BE/ME sorted as well as momentum sorted portfolios. Furthermore they show that conditional skewness is priced and explains cross-sectional variation in expected returns even when factors based on size and book-to-market are included.

[1] Please refer to Appendix B for the VBA code employed to compute equation 2.19, 2.22 & 2.24

Kurtosis in Asset Pricing

The effect of kurtosis is not new to asset pricing literature either but is, however, more recent. Fang and Lai (1997) analyse NYSE returns for the period January 1974 to December 1988 and find kurtosis more prevalent than skewness. This finding prompts the authors to introduce the fourth moment into the unconditional three moment CAPM and develop the concept of cokurtosis or systematic kurtosis. Fang and Lai (1997) present a model where an asset's return is determined by its systematic risk (variance), systematic skewness and systematic kurtosis. Expressed mathematically this is:

$$R^i - R^f = \alpha^i + \beta_i^m\left(R^m - R^f\right) + \beta_i^{m2}\left(R^m - R^f\right)^2 + \beta_i^{m3}\left(R^m - R^f\right)^3 + \varepsilon^i \quad (2.23)$$

Where β_i^{m3} is cokurtosis, calculated as:

$$\beta_i^{m3} = \frac{Cov\left[R^{ei}(R^{em})^3\right]}{E\left[\left(R^{em} - E(R^{em})\right)^4\right]} \quad (2.24)$$

The four moment CAPM expressed in expected return beta form is:

$$E\left(R^i\right) - R^f = \beta_i^m\lambda^m + \beta_i^{m2}\lambda^{m2} + \beta_i^{m3}\lambda^{m3} \quad (2.25)$$

Where the λ's are again factor risk premia. Dittmar (2002) develops the marginal rate of substitution for the four moment CAPM[1]:

$$m_{t+1} = a + bR_{t+1}^m + c(R_{t+1}^m)^2 + d(R_{t+1}^m)^3 \quad (2.26)$$

Where a, b, c and d are functions of Ω_t. Dittmar (2002) extends the quadratic CAPM introducing conditional cokurtosis. Then conditional cokurtosis is (Vorkink, 2003)[2]:

$$\beta_i^k = \frac{E\left(\varepsilon_{i,t+1}^2\varepsilon_{m,t+1}^2\right)}{E\left(\varepsilon_{i,t+1}^2\right)E\left(\varepsilon_{m,t+1}^2\right)} \quad (2.27)$$

Where $\varepsilon_{i,t} = R_{t+1}^{ei} - \alpha_t^i + \beta_{i,t}^m R_t^{em}$. The construction of factors that proxy for the quadratic and cubic terms in Equation 2.23 converts the non-linear asset pricing model into a linear multifactor model.

The cubic CAPM suggests the return of an asset or portfolio is determined by its covariance, coskewness and cokurtosis with the market portfolio. The intuition behind the cubic CAPM, derived from Dittmar's (2002) pricing kernel, is a positive relationship between expected return and cokurtosis.

As with systematic skewness this concept is best explained in the context of investor preferences. Kurtosis is the degree to which, for a given variance, a distribution is weighted toward its tails. Leptokurtic return distributions are characterised by fat tails indicating a larger probability of extreme values (gains or losses). Leptokurtically distributed assets are thus typically viewed as more risky. All

[1] For a detailed derivation of the four moment CAPM's marginal rate of substitution grounded in utility theory, please refer to Dittmar (2002).

[2] Please refer to Appendix B for the VBA code employed to compute equation 2.27

other things being equal, and assuming a risk-averse investor, platykurtically distributed assets should be preferred to leptokurtically distributed assets. Assuming idiosyncratic risk has been diversified away, if an asset then increases a portfolio's kurtosis the risk-averse investor will require a premium to hold the asset. Similarly, one expects assets that decreases a portfolio's kurtosis to have lower expected returns. The cubic CAPM is justified under the argument of a prudent investor's aversion to extreme outcomes.

Dittmar (2002) illustrates that preference restricted nonlinear pricing models are both admissible for the cross-section of returns and are able to significantly improve upon linear single and multifactor models. Conrad, Dittmar, and Ghysels (2013) study the ex-ante skewness and kurtosis implied by option prices and find that ex-ante skewness and kurtosis are indeed related to the cross-section of expected equity returns. Blau, Masud and Whitby's (2013) results are less encouraging. The authors show that stocks with high kurtosis have higher raw returns than stocks with low kurtosis. However, the return premium associated with excess kurtosis excess disappears when controlling for traditional risk factors such as size. This would indicate that kurtosis is not priced in stocks.

2.6 Estimation and Asset Pricing Tests

This section contextualises the estimation methods employed and their associated test statistics. The paper employs a two-pass asset pricing model testing methodology as outlaid by Cochrane (2005). The linear CAPM, higher moment CAPMs and FF3F model is subjected to test statistics that assesses if pricing errors are jointly zero i.e. assess if models accurately characterises the relationship between expected returns and model factors. The first step is to perform time-series regressions for each asset in the sample ($i = 1, \dots, N$):

$$R_t^{ei} = a^i + \beta_i f_t + \varepsilon_t^i \qquad t = 1, \dots, T \tag{2.28}$$

This is followed by cross-sectional regressions of expected returns on the parameters estimated in the time-series regression. For notational purposes it is useful to expresses Equation 2.4 in vector form:

$$E_T(R) = \beta\lambda + \alpha \tag{2.29}$$

The dependent variable $E_T(R)$, is a $N \times 1$ vector of expected portfolio excess returns at time T. β is a $N \times K$ vector of time-series regression coefficients. λ is a $K \times 1$ vector of factor risk premia estimated by cross-sectional regression; and α is a $N \times 1$ vector of pricing errors.

Ordinary Least Squares

Ordinary least squares (OLS) is probably the most widely used class of estimators in finance. Parameters and pricing errors are estimated by[1]:

$$\hat{\lambda} = (\beta'\beta)^{-1}\beta' E_T(R), \qquad \hat{\alpha} = E_T(R) - \hat{\lambda}\,\beta \tag{2.30}$$

[1] Please refer Appendix C for the Python code used to estimate equation 2.30.

In order to test the asset pricing model we need information regarding how parameters and residuals are related to another. The standard OLS formulas for parameter and residual covariance matrices are[1] (Cochrane, 2005):

$$\sigma^2(\hat{\lambda}) = \frac{1}{T}(\beta'\beta)^{-1}\,\beta'\Sigma(\beta'\beta)^{-1} \tag{2.31}$$

$$cov(\hat{\alpha}) = \frac{1}{T}(I - (\beta'\beta)^{-1}\beta')\Sigma(I - (\beta'\beta)^{-1}\beta') \tag{2.32}$$

Where $\Sigma = E_T(\varepsilon\varepsilon')$ and I is a $N \times N$ identity matrix. If asset pricing models adequately characterise the variation in expected returns then $\hat{\boldsymbol{\alpha}}$ should not differ significantly from zero. The standard OLS-constructed Wald statistic that jointly tests the intercepts significance is[1]:

$$J_{OLS} = \hat{\alpha}'cov(\hat{\alpha})^{-1}\hat{\alpha} \sim \chi^2_{N-K} \tag{2.33}$$

If an asset pricing model adequately explains cross-sectional variation in returns, J_{OLS} will be statistically insignificant. Inherent in OLS estimation is the assumption that errors are independent and identically distributed (i.i.d.). It has however been well-documented that errors are not i.i.d., returns have been proven to exhibit serial and cross-sectional correlation (Lo and Mackinlay (1990), Jegadeesh and Titman (1993), Lewellen (2002), amongst others). Time-series correlation is most often addressed by using a long-run covariance matrix while cross-sectional correlation between residuals is most often addressed by using generalised least squares.

Generalised Least Squares

As the OLS cross-sectional regression residuals are correlated with one another, economic theory suggests the application of generalised least squares (GLS). In order to account for this correlation, standard textbook advice suggests using $E(\alpha\alpha') = \frac{1}{T}\Sigma$ as the error covariance[1] (Cochrane, 2005):

$$\hat{\lambda} = (\beta'\Sigma^{-1}\beta)^{-1}\beta'\Sigma^{-1}E_T(R), \qquad \hat{\alpha} = E_T(R) - \hat{\lambda}\,\beta \tag{2.34}$$

Cochrane (2005) does not advocate the use of standard regression formulas to calculate the variance of the above parameters. Equation 2.31, 2.32 and their GLS equivalents assume that the β's are fixed. They are of course not as they are estimated using time-series regressions. Assuming that time-series residuals are i.i.d. over time and independent of factors a Shanken (1992) correction can be used to account for β being estimated:

$$\sigma^2(\hat{\lambda}) = \frac{1}{T}\left[(\beta'\Sigma^{-1}\beta)^{-1}\left(1 + \lambda\Sigma_f^{-1}\lambda\right) + \Sigma_f\right] \tag{2.35}$$

$$cov(\hat{\alpha}) = \frac{1}{T}(\Sigma - (\beta'\Sigma^{-1}\beta)^{-1}\beta')\left(1 + \lambda\Sigma_f^{-1}\lambda\right) \tag{2.36}$$

Where Σ_f is the covariance matrix of the factors used in the time-series regressions. Following Cochrane (2005) we can form the asymptotic GLS test statistic, that corrects for cross-sectional

[1] Please refer Appendix C for the Python code used to estimate equation 2.31 to 2.36.

correlation and the fact that the $\boldsymbol{\beta}$'s are estimated, by dividing pricing errors by their variance covariance matrix[1]:

$$J_{GLS} = T\left(1 + \lambda\Sigma_f^{-1}\lambda\right)\hat{\alpha}'\Sigma^{-1}\hat{\alpha} \sim \chi^2_{N-K} \tag{2.37}$$

Fama-Macbeth Regressions

Fama and Macbeth (1973) pioneered a regression methodology that has become standard practice in asset pricing estimation and testing. The FM is run with factor loading that either remain constant throughout the sample period or alternatively are allowed to vary across time. By keeping betas constant we utilise the full information set available to calculate maximum likelihood cross-sectional estimates.

The full information maximum likelihood FM technique also starts with a set of time-series regressions. In contrast to OLS and GLS however, cross-sectional regressions are run at each point in time[1]:

$$R_t^{ei} = \beta_i\lambda_t + \alpha_t^i \qquad i = 1, \dots, N \tag{2.38}$$

It is important to realise that the dependent variables are the portfolio returns at time t while the independent variables are time-series factor loadings for the entire period. Interestingly α_t^i is the pricing error of portfolio i at point t which means that we have a $T \times N$ vector of pricing errors and a $T \times K$ vector of cross-sectional risk premia. $\hat{\lambda}$ and $\hat{\alpha}$ are then estimated:

$$\hat{\lambda} = \frac{1}{T}\sum_{t=1}^{T} \hat{\lambda}_t \tag{2.39}$$

$$\hat{\alpha} = \frac{1}{T}\sum_{t=1}^{T} \hat{\alpha}_t \tag{2.40}$$

Where $\hat{\boldsymbol{\lambda}}$ is a $1 \times K$ vector of factor risk premia and $\hat{\boldsymbol{\alpha}}$ is a $1 \times N$ vector of pricing errors. Importantly the standard deviations of the cross-sectional regression estimates are used to compute the sampling errors for these estimates by averaging:

$$\sigma^2\left(\hat{\lambda}\right) = \frac{1}{T^2}\sum_{t=1}^{T}\left(\hat{\lambda}_t - \hat{\lambda}\right)^2 \tag{2.41}$$

$$\sigma^2\left(\hat{\alpha}^i\right) = \frac{1}{T^2}\sum_{t=1}^{T}\left(\hat{\alpha}_t^i - \hat{\alpha}^i\right)^2 \tag{2.42}$$

Cochrane (2005) proves that if independent variables do not vary over time the FM regression estimates and errors are equal to the pure cross-sectional OLS regression estimates and errors. Finally the covariance matrix of the sampling errors is calculated as:

$$cov(\hat{\alpha}) = \frac{1}{T^2}\sum_{t=1}^{T} \left(\hat{\alpha}_t - \hat{\alpha}\right)\left(\hat{\alpha}_t - \hat{\alpha}\right)' \tag{2.43}$$

[1] Please refer Appendix C for the Python code used to estimate equation 2.37 to 2.43.

A weighting matrix can be used to account autocorrelation as White (1980) and Newey and West (1987) illustrate. The model can then be tested using a using a Wald statistic[1]:

$$J_{FM} = \hat{\alpha}' cov(\hat{\alpha})^{-1} \hat{\alpha} \sim \chi^2_{N-K} \quad (2.44)$$

Time-varying exposure FM involves constructing β in an alternative manner. β_t in month t is obtained by regressing portfolio returns from months $t - 60$ through $t - 1$. This results in $T - 60$ estimates of β_t vectors with dimensions $N \times K$. The Equation is specified as:

$$R_t^{ei} = \alpha_t^i + \beta_t^i \lambda_t \qquad i = 1, \dots, N \quad (2.45)$$

For $t = 61,62, \dots, T$. In practice the rolling window can be any size but five years of monthly data is customary in asset pricing research. Theoretically the window should be set equal to the length of time that an investor is conditioned on information set (Ω_t). It is important to notice that the cross-sectional regressions decrease in number equal to the length of the rolling window, possibly reducing the overall power of the regressions.

The time-varying FM became very popular exactly because it allows the researcher to discard the assumption that factor exposures remain constant over time. This is important as the central premise of a conditional asset pricing model is that conditional alpha is always zero. Investors are conditioned only by the information set available to them at a specific point in time. By allowing factor exposures to vary over time, FM regressions incorporate only information that condition investors at each point of the time-series. The technique thus allows for the testing of any conditional asset pricing model. The FM test statistic assesses if conditional alpha is always zero and if pricing errors are as a result of time variation in betas.

Research pertaining to the ability of time-variation in beta to explain cross-sectional variation in stock returns is mixed as documented in Section2.2. Pettengill, Sundaram, and Mathur (1995) show that the time-varying beta FM regression is biased against finding a significant relationship between beta and expected return because the relationship between beta and realised returns is conditional on the market return. They therefore argue for the use of constant beta models. Their results are confirmed by Huang and Hueng (2008) who find a positive risk-return relationship in bull markets (i.e. positive market excess returns) and a negative relationship in bear markets (i.e. negative market excess returns).

Generalised Method of Moments

Pure cross-sectional OLS, cross-sectional GLS and FM regressions only efficiently estimates parameters under the assumption of multivariate normal return distributions. This assumption seems unlikely to be true when considering the literate reviewed in Section2.4. Despite the overwhelming evidence of non-normality in returns, most investigations into the risk-return relationship continue to rely on asset pricing tests that assume normality. Estimating unconditional higher moment asset pricing models

[1] Please refer Appendix C for the Python code used to estimate equation 2.44 and 2.45

using regression techniques that assume multivariate normality then seem almost contradictory. Yet, as Cochrane (2005, p.212) states, OLS estimates and errors are "pretty darn good".

A common remedy to the problem of non-normality has been a multivariate adaptation of Hansen's (1982) generalised method of moments. It has the advantage that it can be implemented without having to specify the data gathering process leading to more robust results. Lim (1989) states that the GMM is an appropriate methodology for estimating and testing higher moment models as it avoids the measurement error problem, and also provides asymptotically more efficient estimators by using information from the residual error covariance matrix.

Following Cochrane (2005), the moment conditions used to implement are:

$$g_t(b) = \begin{bmatrix} E(R_t^e - a - \beta f_t) \\ E[(R_t^e - a - \beta f_t) f_t] \\ E(R^e - \beta\lambda) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \tag{2.46}$$

The time-series and cross-sectional estimates are then mapped into a GMM. The parameter vector is (Siriwardane, 2013)[1]:

$$b = \begin{bmatrix} a^1 \\ a^2 \\ \vdots \\ a^N \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_N \\ \lambda^1 \\ \vdots \\ \lambda^K \end{bmatrix} \tag{2.47}$$

b is a $(NK + K) \times 1$ vector of parameters. At this point it is important to note that a^N are the time-series regression intercepts. A $(NK + K) \times T$ vector of sample moment conditions is formed from the parameters as:

$$g_t(b) = \begin{bmatrix} \varepsilon_t^1 \\ \varepsilon_t^2 \\ \vdots \\ \varepsilon_t^N \\ \varepsilon_t^1 f_t \\ \varepsilon_t^2 f_t \\ \vdots \\ \varepsilon_t^N f_t \\ \alpha_t^1 \\ \vdots \\ \alpha_t^N \end{bmatrix} \tag{2.48}$$

[1] Please refer Appendix C for the Python code used to estimate equation 2.47 and 2.48.

Note that α_t^N are the cross-sectional regression residuals while a^N while are the time-series regression intercepts. The spectral density matrix of $g_t(b)$ is formulated as[1]:

$$S = \sum_{j=\infty}^{\infty} E\left(g_t(b) g_{t-j}(b)'\right) \tag{2.49}$$

Hansen (1982) shows that the inverse of the spectral density matrix is the statistically optimal matrix producing estimates with the lowest asymptotic variance. The GMM estimate adjusts b to minimise the quadratic form:

$$\hat{b} = \operatorname{argmin}_b g_T(b)' S^{-1} g_T(b) \tag{2.50}$$

The weighting matrix directs the GMM estimation to emphasize some moments or linear combinations of moments at the expense of others (Cochrane, 2005). The GMM estimate selects a weighting matrix that is statistically optimal in the sense that it favours linear combinations of moments that contain the most information. The standard errors on b are calculated using:

$$\sigma^2\left(\hat{b}\right) = \frac{1}{T} d^{-1} S (d^{-1})' \qquad d = \frac{\partial g_t(b)}{\partial b} \tag{2.51}$$

Where d is the partial derivative used to form a consistent asymptotic variance-covariance matrix. Cochrane (2005) suggests a test of overidentifying restrictions to determine how well the model 'fits' the data. In the case of asset pricing the test statistic determines if pricing errors are too large. The test statistic is:

$$TJ_{GMM} = \left[g_T\left(\hat{b}\right)' S^{-1} g_T\left(\hat{b}\right)\right] \sim \chi^2_{\#moments - \#paramters} \tag{2.52}$$

The test of overidentifying restrictions is fundamentally different from the other test statistics discussed in that it tests the overall fit of the model and not whether the pricing errors are jointly zero.

Harvey and Zhou (1993) find that while GMM methods typically produce more robust test statistics, they predominantly reach the same conclusions as OLS methods. Vorkink (2003) states that GMM provides robustness to asset pricing tests but under non-normality it generally does not lead to fully efficient (minimum variance) estimates and powerful asset pricing tests. He suggest the use of estimators based on the assumption of elliptical symmetry rather than multivariate normality.

HLV Test Statistic

In addressing the issues pertaining to GMM estimation Hodgson, Linton, and Vorkink (2002) utilise a Seeming Unrelated Regression methodology that enables asset pricing model estimation under elliptically symmetric return distributions. In contrast with GMM methods that simply adjusts the standard errors of OLS estimates, the HLV method integrates the elliptical symmetry assumption when constructing coefficients and corresponding standard errors. This is quite a radical departure from typical asset pricing literature. Hodgson, Linton, and Vorkink (2002) show that the resulting estimates are asymptotically efficient resulting in more desirable size and power properties. They suggest that

[1] Please refer Appendix C for the Python code used to estimate equation 2.49 to 2.52.

the HLV methodology allows for sufficiently general distribution assumptions while providing better-behaved test statistics when working with leptokurtic return distributions.

The HLV method incorporates the assumption that stock return distributions are elliptically symmetric. The HLV test is a semi-parametric regression model where the multivariate error density is assumed to be elliptically symmetric. Let $\theta = [\alpha', vec(\beta')']'$ be the $N(K+1)$ vector of parameters in the asset pricing model defined in Equation 2.28. Based on the assumption that the joint distribution of $\{R_t^e, f_t\}_{t=1}^T$ is elliptical, the HLV estimate (θ) has the following asymptotic distribution:

$$\sqrt{\theta}\left(\tilde{\theta} - \theta_0\right) \Rightarrow N(0, I^{-1}) \tag{2.53}$$

Where $\Rightarrow$ denotes weak convergence of probability measures, θ_0 is the true value of θ, and I is Fisher's asymptotic information matrix. Given the estimates of parameters $\tilde{\theta}$, Wald tests can be constructed in the usual manner. Thus the HLV test statistic:

$$J_{HLV} = \hat{\alpha}' cov(\hat{\alpha})^{-1} \hat{\alpha} \sim \chi^2_{N-K} \tag{2.54}$$

In this study an attempt was made to replicate the HLV estimation and testing techniques as applied to higher moment asset pricing models in Vorkink (2003). The results were not meaningful as the stopping criterion in the implementation of the Fisher scoring algorithm used to maximise the likelihood had to be very weak to ensure convergence (typically $\frac{\delta L}{L} \approx 1\%$ where L is the likelihood). Moreover we found that the parameter vector could vary by as much as 10% for a corresponding 1% variation in the likelihood.

This study does not claim that the HLV method does not work, in fact it is very possible that we implemented it incorrectly and we didn't try to contact the authors directly. However it must be pointed out that the code which was made publically available and the formulae given in Vorkink (2003) only hold for models in which the number of portfolios is the same as the number of factors. For these reasons it was decided to focus on the OLS, GLS, FM and GMM estimation and testing in this study.

3. Data

The data utilised in the study is from the FinData@Wits database. Total returns, market capitalisation, book to market equity, and zero trading days were sourced for all firms listed on the JSE on a monthly basis for the period January 1992 to December 2013. The total return on a share is calculated as follows:

$$R_t^i = \frac{P_{t+1}^i + D_{t+1}^i}{P_t^i} - 1 \qquad (3.1)$$

Where P is price and D is dividend. The FinData@Wits database is constructed utilising a number of different data sources. The majority of the data is sourced from I-Net Bridge and McGregor BFA while JSE monthly bulletins were used to account for corporate actions. Survivorship bias often impacts the results of firm level empirical studies in economic literature. The FinData@Wits accounts for survivorship bias by utilising a complete sample that includes delisted and suspended shares. The sample used in this study as such consists of 1384 shares that are/were listed on the JSE during some point between January 1992 and December 2013. Delisted or suspended shares are allocated a zero return.

FinData@Wits further provided total returns for the FTSE/JSE All-Share Index (J203) for period January 1992 to December 2013 to serve as the proxy for the market portfolio. The J203 was selected to enable comparison with previous studies completed using JSE data. The monthly yield on the three month treasury bills were obtained directly from the South African Reserve Bank for the same period to proxy for the risk free rate.

Portfolios are formed using 12 month averages of sorting criteria (e.g., market equity and BE / ME) to provide a holistic view of what occurred during the year and mitigate the effect of extreme market movements between December and January each year. This constraint results in portfolios with returns commencing in January 1993 and ending December 2013. Portfolio as such have 252 monthly data points.

The FTSE / JSE Industry Classification Benchmark (ICB) consists of four levels of classification and at the top level there are ten Industries. The constituents of the All Share Index (J203) are classified by the ten industries. The industry indices include constituents of the All Share Index (J203) that are classified in the industry after which the index is named. The indices are:

Table 3.1 FTSE / JSE industry classification indices

Index Code	**Index Name**
J500	Oil & Gas
J510	Basic Materials
J520	Industrials
J530	Consumer Goods
J540	Health Care
J550	Consumer Services
J560	Telecommunication
J570	Utilities
J580	Financials

J590	Technology

Data was extracted from INET BFA for FTSE / JSE industry indices for the period January 1997 to December 2013. The oil & gas industry index (J500) and the utilities industry index (J570) were excluded from the sample as oil & gas industry index (J500) data only starts from February 2006 and no data are available for the utilities industry (J570). The remaining eight indices proxy for industry portfolios.

4. Methodology

The primary purpose of the study is to provide clarity on the relationship between higher moment asset pricing models and the cross-section of average returns on the JSE. The methodology is structured to answer the following hypotheses:

H_0: Sample portfolio returns are normally distributed

H_1: Coskewness is not priced on the JSE

H_2: Cokurtosis is not priced on the JSE

H_3: Model pricing errors are jointly zero

These hypotheses are defined more clearly later on in the study. Questions further addressed in a less formal manner are:

- Are the size and value effects present in the data? How are they related to the higher moments?
- Do unconditional or conditional models fit the data best?
- What is the impact of the distributional nature of data on the estimation method used within an asset pricing context?

Broadly the methodology of the paper consists of: (1) univariate and multivariate distributional analysis of sample factors and portfolios; and (2) the estimation and testing of the CAPM, FF3F, quadratic CAPM and cubic CAPM using OLS, GLS and GMM techniques for sample portfolios. The ME sorted and BE / ME sorted portfolio sets are selected as sample data as they are often used in empirical asset pricing research and as such results are directly comparable to a large body of literature.

4.1 Portfolio Sorts and Factor Construction

This sections describes the methodology employed in the construction of sample portfolios and portfolios that proxy for model risk factors.

In addressing illiquidity or thin trading constraints the study subjects shares to a liquidity filter prior to portfolio formation. McClelland, Auret and Wright (2014) show the choice of an appropriate thin-trading filter is a function of the choice of the beta. The authors suggest a simple adjusted OLS model when not working with daily data. The adjusted OLS model adjusts for the amount of days traded. In

order to compile the liquidity filter, zero days trading[1], shares are evaluated each day within the sample period to determine trading volume. If no trading occurred then shares are assigned a one for the day, or alternatively, if trading took place shares are assigned a zero. Zero days trading is then the summation of the binary liquidity variable over the desired time horizon e.g. monthly, quarterly or annually. Stocks were only included in portfolios should their zero days trading not exceed 150 days for the preceding 12 months.

Size and Value Portfolios

Equally-weighted and value-weighted portfolios are formed by independent decile sorts on market equity (size) and BE/ME (value). Portfolio sorts occur annually in January for each year between 1993 and 2013. The 12 month averages of market equity and BE / ME are used as an indicators rather than the end-of-year value so as to provide a holistic view of what occurred during the year and mitigate the effect of extreme market movements[2].

Size-value sorted portfolios are formed by dependent quintile sorts over the same time period. First portfolios are sorted on average market equity into quintile portfolios. Each size quintile is subsequently sorted into five average BE/ME portfolios. The result is 25 size-value sorted portfolios. Size-value weighted portfolios are only compiled on a value weighted basis in order to keep results manageable.

Industry Indices

The FF3F model uses proxies for size and value effects as factors. The use of size and value portfolios when estimating the F3FF model could lead to linear dependence between sample portfolios and factors. Industry indices from the FTSE/JSE Africa Index series are used as a proxy for industry portfolios to provide an independent point of comparison.

FF3F Factors

Fama and French (1993) construct their factors by first forming portfolios sorted on size and value, independent sorts of two size portfolios and three value portfolios to be more precise. Six portfolios are formed on the breakpoints of size and value sorts. The SMB factor is constructed by subtracting the simple average of the returns of the big-stock portfolios (Big / Low, Big / Medium, Big / High) from the simple average of the returns of the small-stock portfolios (Small / Low, Small / Medium, Small / High). Similarly, Fama and French (1993) construct the HML factor by subtracting the simple average of the returns on the two high-BE/ME portfolios (Small / High, Big / High) from the average of the returns on the two low- BE/ME portfolios (Small / Low, Big / Low). Factors are formed in this manner to ensure limited correlation.

[1] We would like to thank the University of Witwatersrand for providing monthly zero days trading data.
[2] Please refer to Appendix A for the Microsoft SQL Server code used to automate portfolio and factor formation

Due to a limited universe of shares listed on the JSE, size and value factors are constructed based on univariate quintile sorts. The SMB factor is equivalent to simultaneously holding a long position in a portfolio of the lowest market equity quintile stocks (small portfolio) and a short position in a portfolio of the highest market equity quintile stocks (big portfolio). Similarly the HML factor is equivalent to holding a long position in a portfolio that consists of the highest BE/ME quintile stocks (value portfolio) and short position in a portfolio of the lowest BE/ME quintile stocks (growth portfolio).

Conditional Skewness and Conditional Kurtosis Factors

Following Harvey and Siddique (2000) and Vorkink (2003) conditional coskewness and cokurtosis factors are constructed. The process is initiated by calculating conditional coskewness and cokurtosis for all firms within the sample using Equations 2.21 and 2.26. Coskewness and cokurtosis in month t are obtained by computing return data from months $t - 60$ through $t - 1$.

Three portfolios are formed using univariate sorts by ranking firms based on coskewness (cokurtosis) in the following manner: the lower 30 percent forms portfolio S^- (K^-); the middle 40 percent forms portfolio S^0 (K^0); and the upper 30 percent forms portfolio S^+ (K^+). As a 60 month window is selected to compute conditional coskewness and cokurtosis portfolio sorts occur annually in January for each year between 1997 and 2013. The resultant portfolios having 204 monthly data points.

Harvey and Siddique (2000) let the excess return on portfolio S^- proxy for the conditional coskewness factor while Vorkink (2003) lets the excess return on K^+ proxy for the conditional cokurtosis factor. Harvey and Siddique (2000) presents an interesting alternative to constructing the conditional coskewness and conditional cokurtosis factors. The technique is similar to that used to construct the FF3F model proxy portfolios. The conditional coskewness factor is equivalent to simultaneously holding a long position in portfolio S^- and a short position in portfolio S^+. Similarly the conditional cokurtosis factor is equivalent to simultaneously holding a long position in portfolio K^+ and short position in portfolio K^-. The conditional quadratic and cubic CAPM models are estimated by regressing sample portfolio returns on market, conditional skewness and conditional kurtosis factors.

We expect an excess return on the S^- portfolio due to investor non-increasing absolute risk aversion. Harvey and Siddique (2000) document the average annualised spread between the returns on the S^- and S^+ portfolios at 3.60 percent over the period July 1963 to December 1993. Similarly we expected an excess return on the K^+ portfolio due to investor aversion to extreme outcomes.

4.2 Distribution Analysis

Return distributions of size and value sorted portfolios as well as proxy industry portfolios (FTSE/JSE Africa Index Series) are assessed using various statistics to test normality and linearity.

Individual Portfolios

Measures of location, dispersion and shape are calculated for sample portfolios. The measures calculated for both equally-weighted and value-weighted portfolios include:

(a) Mean

(b) Standard deviation
(c) Skewness
(d) Unconditional coskewness
(e) Kurtosis
(f) Unconditional cokurtosis

Mean, unconditional systematic skewness and unconditional systematic kurtosis are tested to see if they are significantly different from zero.

H_0 is formally defined as:

$H_0: R_i \sim N(.)$

Portfolio skewness and kurtosis are assessed to determine whether they are consistent with **H_0**. Finally a goodness-of-fit test, the Jarque–Bera (1987) test statistic, is compiled. J_{JB} tests whether sample data have the skewness and kurtosis matching a normal distribution:

$$J_{JB} = \frac{T}{6}\left(Skewness^2 + \frac{1}{4}(Kurtosis - 3)^2\right) \sim \chi_2^2 \quad (4.1)$$

The J_{JB} statistic asymptotically has a chi-squared distribution with two degrees of freedom and tests the hypothesis that the portfolio returns are normally distributed. We are therefore able to reject **H_0** if J_{JB} is statically significant.

Multivariate Normality

One of the assumptions underlying the CAPM is that asset returns follow a joint normal distribution. The joint distributions of size-sorted, value-sorted, size-value sorted and industry portfolios are assessed using various multivariate tests of normality:

(a) Multivariate Shapiro-Wilk (Royston, 1983) test
(b) Henze-Zirkler (1990) test
(c) Mardia's skewness and kurtosis test (Mardia, 1970)
(d) Adjusted Mardia's skewness test
(e) Doornik-Hansen omnibus test (Doornik and Hansen, 2008)

The calculation of the above test statistics are quite complex and was completed with an automated add-in for Eviews. A detailed description of their computation and interpretation can be found in the cited sources, herein follows only a brief note on their interpretation as relevant to this study. The Multivariate Shapiro-Wilk (Royston, 1983) test follows an asymptotically chi-squared distribution with N degrees of freedom and tests the hypothesis that the sample portfolio returns are jointly normally distributed. The Henze-Zirkler (1990) test is interpreted by a Wald test statistic for multivariate normality on the log of J_{HZ}. **H_0** is rejected if J_{HZ} is statistically significant. Mardia's multivariate skewness is asymptotically chi-squared distributed with $N(N+1)(N+2)/6$ degrees of freedom. Mardia's multivariate kurtosis is asymptotically normally distributed with mean $N(N+2)$ and variance $8N(N+2)/T$. Finally the Doornik-Hansen (2008) test is asymptotically chi-squared

distributed with N degrees of freedom and tests the hypothesis that the sample portfolio returns follow a joint normal distribution.

Linearity

Fama-Macbeth (1973) assess the linearity of stock returns by hypothesis $E(\hat{\lambda}_2) = 0$ from the regression equation:

$$E_T(R^i) = \lambda_{0,t} + \lambda_{1,t}\beta_i + \lambda_{2,t}\beta_i^2 + \lambda_{3,t}\sigma(\varepsilon_i) + \mu_{i,t} \quad (4.2)$$

Where $\sigma(\varepsilon_i)$ is the standard deviation of time-series regression residual for asset i and $\mu_{i,t}$ is the FM regression residual for asset i. They find that the linearity hypothesis cannot be rejected.

Similarly we complete a Ramsey (1969) Regression Specification Error test to assess if a linear model fits return data. Inherent in checking if the quadratic and cubic term are priced, is a test of linearity. Thus this study does not formally present a hypothesis of linearity. If **H_1** and **H_2** are incorrect then:

$$\hat{\lambda}^{m2} = \hat{\lambda}^{m3} = 0$$

From the FM regression equation:

$$E_T(R^{ei}) = \lambda^m\beta_i^m + \lambda^{m2}\beta_i^{m2} + \lambda^{m3}\beta_i^{m3} + \alpha^i \qquad i = 1, \dots, N \quad (4.3)$$

Linearity is thus rejected if either **H_1** or **H_2** are rejected.

4.3 Estimation

The CAPM, quadratic CAPM, cubic CAPM and FF3F models are estimated by OLS, GLS, FM and GMM techniques. This section provides a comprehensive description of the estimation techniques employed as well as the details regarding the manner in which equations are specified.

Time-series Regressions

Fama and French (1993) run time-series regressions of their 25 size-values portfolios on CAPM and FF3F model factors to evaluate the models' ability to explain variation in returns. Similarly we complete time-series regressions of each of the sample 25 size-value sorted portfolios on CAPM, quadratic CAPM, cubic CAPM and FF3F models:

$$R_i^e = a^i + \beta_i \boldsymbol{f_t} + \varepsilon_t^i \qquad t = 1, \dots, T \quad (4.4)$$

Where $\boldsymbol{f}$ is a $T \times K$ vector of factors. The moments that map time-series OLS into a GMM are:

$$\begin{bmatrix} E(R_t^i - a - \beta f_t) \\ E[(R_t^i - a - \beta f_t)f_t] \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad (4.5)$$

A Newey-West (1987) weighting matrix is used to ensure standard errors are robust to autocorrelation and heteroskedasticity.

Cross-sectional Regressions

Expected portfolio returns are regressed on parameters ($\boldsymbol{\beta}$) estimated in the time-series regressions. First we specify cross-sectional regression equation:

$$E_T(R^{ei}) = \beta_i^m \lambda^m + \beta_i^{Skew} \lambda^{Skew} + \alpha^i \qquad i = 1, \dots, N \qquad (4.6)$$

Then **H_1** implies that:

$$H_2: \lambda^{Skew} = 0$$

We can reject **H_1** if λ^{Skew} is significantly different from zero. Note that β_i^{Skew} can be derived from the market factor squared (the unconditional quadratic CAPM) or from portfolio S^- (the conditional quadratic CAPM). The same approach is applied to test **H_2**, we run regressions:

$$E_T(R^{ei}) = \beta_i^m \lambda^m + \beta_i^{Skew} \lambda^{Skew} + \beta_i^{Kurt} \lambda^{Kurt} + \alpha^i \qquad i = 1, \dots, N \qquad (4.7)$$

H_2 implies that:

$$H_2: \lambda^{Kurt} = 0$$

We reject **H_2** if λ^{Kurt} is statistically significant. It should be noted here that β_i^{Kurt} can be derived from the market factor cubed (the unconditional cubic CAPM) or from portfolio K^+ (the conditional cubic CAPM). The CAPM and FF3F models are estimated for comparative purposes by running cross-sectional regressions:

$$E_T(R^{ei}) = \beta_i^m \lambda^m + \alpha^i \qquad i = 1, \dots, N \qquad (4.8)$$

$$E_T(R^{ei}) = \beta_i^m \lambda^m + \beta_i^s \lambda^s + \beta_i^v \lambda^v + \alpha^i \qquad i = 1, \dots, N \qquad (4.9)$$

Cross-sectional regressions are executed using OLS, GLS, FM and GMM techniques to assess the impact of estimation methodology on our hypotheses. Furthermore FM regressions are run with constant and rolling betas to assess the impact time-variation has on our hypotheses.

4.4 Asset Pricing Model Tests

The central premise of any asset pricing model is that factors explain the cross-sectional variation in expected return. If asset pricing models adequately characterise the relationship between risk and expected return then cross-sectional pricing errors ($\boldsymbol{\alpha}$) should be jointly zero. **H_3** therefore is:

$$H_3: \boldsymbol{\alpha} = 0$$

We test this **H_3** on size, value and industry portfolios with the test statistics detailed in Section 2, recalling Equations 2.33, 2.37, 2.44 and 2.52:

$$J_{OLS} = \hat{\boldsymbol{\alpha}}' cov(\hat{\boldsymbol{\alpha}})^{-1} \hat{\boldsymbol{\alpha}} \sim \chi^2_{N-K} \qquad (2.33)$$

$$J_{GLS} = T\left(\mathbf{1} + \boldsymbol{\lambda}\boldsymbol{\Sigma}_{\boldsymbol{f}}^{-\mathbf{1}}\boldsymbol{\lambda}\right)\hat{\boldsymbol{\alpha}}'\boldsymbol{\Sigma}^{-\mathbf{1}}\hat{\boldsymbol{\alpha}} \sim \chi^2_{N-K} \qquad (2.37)$$

$$J_{FM} = \hat{\boldsymbol{\alpha}}' cov(\hat{\boldsymbol{\alpha}})^{-1} \hat{\boldsymbol{\alpha}} \sim \chi^2_{N-K} \qquad (2.44)$$

$$TJ_{GMM} = \left[g_T(\widehat{\boldsymbol{b}})' S^{-1} g_T(\widehat{\boldsymbol{b}})\right] \sim \chi^2_{\#moments - \#parameters} \quad (2.52)$$

All models follow a chi-squared distribution with $N - K$ degrees of freedom except the J_{GMM} where degrees of freedom equal the number of parameters in the GMM subtracted form the number of moment conditions.

4.5 Simulations

Finally we perform a Monte Carlo experiment to assess certain characteristics of the chosen test statistics. In particular we utilise the graphical methods developed by Davidson and MacKinnon (1998, hereafter DM) to investigate the size and power properties of our hypothesis tests. This is meant to determine the suitability of the above mentioned estimation techniques when working with non-normal data. The DM methodology is highly informative, comparatively simple to implement and it yields easily interpretable graphs. The graphs are based on the empirical p-value distribution of the given test statistics. The p-value (P) of the test statistics employed in this paper is the probability that the model pricing errors are jointly zero (i.e. H_3: $\alpha = 0$) or alternatively the overall fit of the model in the case of J_{GMM} . At any point x_i falls between zero and one and is defined by:

$$\hat{F}(x_i) = \frac{1}{S}\sum_{j=1}^{S} I\,(p_j < x_i) \quad (4.10)$$

Where S is the number of simulations and $I\,(p_j < x_i)$ is an indicator function that takes a value of 1 if the argument is true and zero otherwise. Although possible to evaluate $\hat{F}(x_i)$ at every data point, DM suggest that choosing m points ($x_i, i = 1, \dots, m$) in a manner that provides an accurate depiction of the $(0, 1)$ interval. For the purposes of this paper x_i is equal to:

$$x_i = 0.001, 0.002, \dots, 0.999 \quad (m = 1000) \quad (4.11)$$

This selection ensures that plotted lines are smooth and not jagged in the graphical techniques utilised.

The exercise commences by simulating an artificial dataset of returns**,** $\tilde{\boldsymbol{r}}$, a $T \times N$ matrix. Here T represents the size of the sample (length of the time-series) and N the number of portfolios. The simulation follows a factor model similar to equation 2.4 such that:

$$\tilde{\boldsymbol{r}} = \boldsymbol{\alpha} + \widehat{\boldsymbol{\beta}}\boldsymbol{f}' + \tilde{\boldsymbol{\varepsilon}} \quad (4.12)$$

Where $\boldsymbol{\alpha}$ is a $N \times 1$ vector that takes the value of 0 or 0.01 and $\widehat{\boldsymbol{\beta}}$ is the $N \times K$ vector of time-series coefficients identified in Section5.2, K being the number of factors. $\boldsymbol{f}$ is the $T \times K$ vector of the FF3F factors constructed in Section5.1, i.e. r_M, SMB and HML. $\tilde{\boldsymbol{\varepsilon}}$ is a $N \times T$ vector of error terms that follow a predefined distribution.

For each simulation we construct a series of null returns, $\tilde{\boldsymbol{r}}^{\boldsymbol{n}}$**,** and a series of alternative returns, $\tilde{\boldsymbol{r}}^{\boldsymbol{a}}$ in the following manner. The calculation starts by finding the product of $\widehat{\boldsymbol{\beta}}$ and $\boldsymbol{f}$**.** This is followed by the addition of a randomly sampled vector of residuals $\tilde{\boldsymbol{\varepsilon}}$ drawn from a predefined distribution. Errors are drawn from both the central $t_{(3)}$ and central $t_{(5)}$ distributions to investigate the effect of varying degrees of kurtosis on the selected test statistics. Residuals are drawn from the $\chi^2_{(4)}$ to investigate the effect

of skewness on the test statistics and finally samples are drawn from a normal distribution for comparison. In all cases $E(\tilde{\boldsymbol{\varepsilon}}) = 0$ while $Var(\tilde{\boldsymbol{\varepsilon}})$ is set equal to $Var(\hat{\varepsilon})$ calculated in Section5.3.

Next we proceed to estimate the FF3F model using OLS, GLS, FMB and GMM techniques. Using the parameters from a given estimation, we calculate test statistics J_{OLS}, J_{GLS}, J_{FM} and J_{GMM} and store the corresponding p-values. This process is repeated 10 000 times ($S = 10000$).

The Monte Carlo methodology thus investigates the sensitivity of each of the tests to different degrees of kurtosis and asymmetry by generating artificial returns with different noise models. This enables the analysis of size to power characteristics of each test statistic and determine which is best suited to evaluate asset pricing models within a South African context[1].

Size properties of the test statistics are evaluated using DM p-value plots for each distribution. The construction of the p-value plot involves graphing x_i on the x-axis and $\hat{F}(x_i)$ on the y-axis for all test statistics under a specific distribution. Test statistics that generate accurate p-values under the null should be distributed as uniform $(0, 1)$, the p-value plot should therefore be close to 45° line. DM (1998) show that if the p-value plot is above (below) the 45° line, the test-statistic systematically over (under) rejects the null.

Finally power properties are evaluated by constructing DM size-power curves. $\hat{F}^n(x_i)$ is plotted on the x-axis while $\hat{F}^a(x_i)$ is plotted on the y-axis. DM show that size problems can be removed from the analysis by plotting $\hat{F}^n(x_i)$ rather than x_i on the x-axis allowing a comparison of power properties. Test statistics have greater power, that is they correctly reject the alternative more often, if their size-power curves are above the size-power curves of the other test statistics that were assessed in the simulation.

[1] Please refer to Appendix D for the Python code employed to complete the Monte Carlo experiment

5. Empirical Results

We use two sample periods within our empirical analysis: (1) January 1993 to December 2013; and (2) January 1997 to December 2013. A longer sample period is preferred but we are forced to use January 1997 and December 2013 period for two reasons: (1) conditional higher moment models require a 60 month time-window in estimating coskewness and cokurtosis; and (2) return data for industry portfolios are only available from January 1997. If results were constant over differing sample periods, one could constrain the empirical analysis to sample period January 1997 and December 2013 but specific characteristics of the preceding four years dictate their inclusion to support certain key findings.

This section commences with a periodic analysis of sample factors followed by a distributional analysis of sample portfolios. Then we estimate and test the CAPM, quadratic CAPM, cubic CAPM and FF3F using time-series and cross-sectional regressions techniques. It concludes with the Monte Carlo experiment designed to derive the size and power characteristics of the test statistics employed in this work.

5.1 Factor Analysis

The factors by construction estimate risk premia e.g. the market or the size premium. In assessing whether premia are concentrated in a specific sub-period factors are analysed over sequential sub-periods to determine dispersion. Furthermore we assess correlation between factors as they are constructed form the same universe of shares which could lead to multicollinearity.

Size and Value Premia

Even though it is not the central focus of the study, an analysis of the size and value premia produces interesting results. The size and value factors are constructed in a near identical manner in which Basiewicz and Auret (2009) complete univariate portfolio sorts to determine the magnitude of size and value effects. Recall that Basiewicz and Auret (2009) document the size and value premia during the period July 1992 to July 2005 for value weighted portfolios at 0.87 percent per month and 1.5 percent per month, respectively. On equally weighted portfolios they document the size and value effects at 1.1 percent per month and 1.56 percent per month, respectively.

On value weighted portfolios for the period January 1993 to December 2013 we document the size and value premia at 0.74 percent and 0.73 percent per month respectively (refer to Table 5.1). On equally weighted portfolios for the same period we document the size and value premia at 1.0 percent and 1.46 percent per month. Value weighted factor mean returns are significantly different from zero at the ten percent level. Equally weighted size and value premium are statistically significant at the five percent and one percent level, respectively. Daniel and Titman (1999) show that value-weighting decreases the impact of trading costs. This might explain why equally weighted portfolios outperform value weighted portfolios. This study thus confirms the results of Van Rensburg and Robertson (2003) and Basiewicz and Auret (2009) based on the January 1993 to December 2013 sample period.

However, the magnitudes of size and value premia are more in line with the findings of Basiewicz and Auret (2009).

Table 5.1 Periodic analysis of factor returns

Market, Size, Value, "SKEW" and "KURT" factors are constructed. The Market factor is the excess return over the risk-free rate of the FTSE/JSE All Share Index. The size or SMB factor is equivalent to simultaneously holding a long position in a portfolio of the lowest market equity quintile stocks (small portfolio) and a short position in a portfolio of the highest market equity quintile stocks (big portfolio). Similarly the value or HML factor is equivalent to holding a long position in a portfolio that consists of the highest BE/ME quintile stocks (value portfolio) and short position in a portfolio of the lowest BE/ME quintile stocks (growth portfolio). The conditional coskewness factor is equivalent to simultaneously holding a long position in portfolio S^- and a short position in portfolio S^+. Similarly the conditional cokurtosis factor is equivalent to simultaneously holding a long position in portfolio K^+ and short position in portfolio K^-. Size, Value, "SKEW" and "KURT" premia are calculated for sub periods: January 1993 to December 1996; January 1997 to December 2003; January 2004 to December 2008; January 2009 to December 2013. ***, **, * refer to statistical significance the 1, 5 and 10% levels, respectively.

Value Weighted Portfolios

Sample period	Market	SMB	HML	SKEW	KURT
January 1993 to December 2013	0.707% **	0.742% *	0.731% *		
January 1997 to December 2013	0.683% *	0.082%	0.347%	-0.970% **	-0.248%
Sub-periods					
January 1993 to December 1996	0.823%	3.609%	2.401%		
January 1997 to December 2003	0.190%	0.281%	0.831%	-1.113%	0.093%
January 2004 to December 2008	0.930%	0.242%	0.803%	-0.901%	-0.406%
January 2009 to December 2013	1.153%	-0.354%	-0.777%	-0.871%	-0.578%

Equally weighted portfolio

Sample period	Market	SMB	HML	Skew	Kurt
January 1993 to December 2013	0.707% **	0.999% **	1.457% ***		
January 1997 to December 2013	0.683% *	0.140%	0.707% **	-0.373%	-0.135%
Sub-periods					
January 1993 to December 1996	0.823%	4.731%	4.724%		
January 1997 to December 2003	0.190%	0.049%	1.040%	-0.491%	0.093%
January 2004 to December 2008	0.930%	0.431%	0.808%	-0.144%	-0.406%

January 2009 to December 2013	1.153%	-0.019%	0.163%	-0.448%	-0.578%

The most interesting result related to FF3F factors is apparent in the shorter sample period. On value weighted portfolios for the period January 1997 to December 2013 we document the size and value premia at 0.08 percent and 0.35 percent per month respectively (refer to Table 5.1). On equally weighted portfolios for the same period we document size and value premia at 0.14 percent and 0.71 percent per month respectively. Value weighted size and value premia and the equally weighted size premium are no longer significantly different from zero. Only the equally weighted value premium is still statistically significant at the five percent. Excluding returns between January 1993 and December 1996 therefore results in the size and value premia disappearing.

A periodic analysis confirms that the size and value premia respectively, for the period January 1993 to December 1996, are 3.61 percent and 2.4 percent per month on a value weighted basis, and 4.73 percent and 4.72 percent per month on an equally weighted basis. Comparatively, the size and value premia respectively, for the period January 1997 to December 2008, are 0.26 percent and 0.82 percent per month on a value weighted basis, and 0.24 percent and 0.92 percent per month on an equally weighted basis. Finally, and most interestingly, the size and value premia reverse for the period January 2009 to December 2013. The "big" portfolio out performs the "small" portfolio by 0.35 percent per month on a value weighted basis and 0.02 percent on an equally weighted basis while the "growth" portfolio outperforms the "value" portfolio by 0.78 percent on a value weighted basis. The low value portfolio does not outperform the high value portfolio on an equally weighted basis which explains why only the equally weighted value premium is statistically significant for the period January 1997 to December 2013. This result confirms a reversal in the size and value premia on the JSE over the most recent five years.

There have been many explanations relating the Fama and French (1992) risk factor proxies to economic theory. The most common is that risk factors capture financial distress or the illiquidity of smaller firms. Investors are rewarded for investing in more risky firms, i.e. small firms with relatively "low value". Hence the size and value premium. Attributing the reversal of the size and value premia in the most recent five years of the sample period to economic theory is more complex. Investors could have realised profits by investing in large companies with "high values", companies that are viewed by the investment community as relatively "safe". A reducing or reversing size effect is not new to the literature. Cochrane (1999) shows the reversal of the size premium in the U.S. while Dimson, Marsh, and Staunton (2002) conclude that the size effect is reversing at a global level based on their analysis of 19 markets worldwide. Strugnell, Gilbert and Kruger (2011) find tentative evidence that the size effect is reducing over time. They evaluate JSE return data for the period January 1994 to October 2007 and graphically illustrate that the size premium is shrinking over time. More specifically they authors show that it follows a downward linear trend.

Gompers and Metrick (2001) attributed the reversal of the size premium to the increased amount of institutional investors that either have a preference for or is mandated to invest in large capitalization stocks. They argue that the increasing demand for "big" stocks is driving the increased prices and

thereby increasing returns. Strugnell *et al.* (2011) relate the reducing size premium to the hypothesis that the size premium historically has in part represented market inefficiency. The idea is that this statistical arbitrage opportunity has been closed out over time by the participation of greater expertise in South African financial markets, increased asset management competition and access to quicker and more accurate information.

A Coskewness Premium

Coskewness factors can be constructed in similar manner to the FF3F factors. Harvey and Siddique (2000) hypothesise a negative relationship between conditional coskewness and expected return. If their hypothesis is correct then a spread should exist between the expected returns on S^- and S^+ portfolios. Harvey and Siddique (2000) document the average annualised spread between the returns on the S^- and S^+ portfolios at 3.60 percent over the period July 1963 to December 1993. Theoretically they base this relationship on investor non-increasing absolute risk aversion.

We compile portfolios using Harvey and Siddique's (2000) methodology and document the spread between the S^- and S^+ portfolios for the period January 1997 to December 2013 at -0.97 percent per month on a value weighted basis and -0.37 percent on an equally weighted basis (refer to Table 5.1). The mean of the value weighted spread is statistically significant at the five percent level and remains reasonably constant over the sub periods within the periodic analysis. This means that an arbitrage portfolio of a long position in S^+ and a short position in S^- would have earned an investor an annualised return of 12.28 percent over the 16 years of the sample period. Even though we do not find support for Harvey and Siddique's (2000) hypothesis of a negative relationship between conditional coskewness and expected return, this is still an extra ordinary result. The "coskewness" premium is larger than the size and value premia over the January 1997 to December 2013 sample period. While this does not definitively contradict the negative relationship between coskewness and expected return, it most assuredly cast doubt on the validity of the hypothesis that skewness enters asset pricing on the basis of investor non-increasing absolute risk aversion.

Dittmar (2002) justifies a cubic pricing kernel under the argument of investor aversion to extreme outcomes and posits a positive relationship between cokurtosis and expected return. We document the spread between the K^+ and K^- for the period January 1997 to December 2013 at -0.25 percent per month on a value weighted basis and -0.14 percent per month on an equally weighted basis (refer to Table 5.1). Unfortunately neither estimate is significantly different from zero so no reliable conclusions can be drawn.

Table 5.2 Factor Correlation Matrices

JSE listed share returns are sorted into portfolios for two sample periods: (1) January 1993 to December 2013; and (2) January 1997 to December 2013. The size or SMB factor is equivalent to simultaneously holding a long position in a portfolio of the lowest market equity quintile stocks (small portfolio) and a short position in a portfolio of the highest market equity quintile stocks (big portfolio). Similarly the value or HML factor is equivalent to holding a long position in a portfolio that consists of the highest BE/ME quintile stocks (value portfolio) and short position in a portfolio of the lowest BE/ME quintile stocks (growth portfolio). The conditional coskewness factor is equivalent to simultaneously holding a long position in portfolio S^- and a short position in portfolio S+. Similarly the conditional cokurtosis factor is equivalent to simultaneously holding a long position in portfolio K^+ and short position in portfolio K^-. Table 5.2 presents the correlation matrices of constructed factors.

January 1993 to December 2013

	Value-weighted factors		
	Market	SMB	HML
Market	1	-0.350	-0.223
SMB	-0.3505	1	0.337
HML	-0.223	0.337	1

	Equally-weighted factors		
	Market	SMB	HML
Market	1	-0.177	-0.061
SMB	-0.177	1	0.715
HML	-0.061	0.715	1

January 1997 to December 2013

	Value-weighted factors				
	Market	SMB	HML	SKEW	KURT
Market	1	-0.511	-0.318	-0.210	0.129
SMB	-0.511	1	0.281	0.256	-0.154
HML	-0.318	0.281	1	-0.058	-0.058
SKEW	-0.210	0.256	-0.058	1	-0.220
KURT	0.129	-0.154	-0.058	-0.220	1

	Equally-weighted factors				
	Market	SMB	HML	SKEW	KURT
Market	1	-0.377	-0.281	-0.060	0.106
SMB	-0.377	1	0.482	-0.036	-0.144
HML	-0.280	0.482	1	-0.075	-0.088
SKEW	-0.060	-0.036	-0.075	1	-0.280
KURT	0.106	-0.144	-0.088	-0.280	1

Factor Correlation

Multicollinearity refers to the phenomenon where two or more predictor variables in a multiple regression model are highly correlated i.e. variables are linearly dependent on one another. This does not impact the overall ability of the regression to explain variation in the dependent variable but does effect individual coefficient estimates. For example, a factor might sufficiently explain variation in the dependent variable until other independent variables are added to the regression resulting in the original factor coefficient no longer being significant. As the purpose of asset pricing is to accurately identify causal relationships between asset returns and independent variables, one should avoid using highly correlated variables.

Table 5.2 lists the correlation between the factors used in the regression analysis. Value-weighted factor correlation for the period January 1993 – December 2013 appears manageable. The highest correlation is between SMB and market factors documented at -0.35. Due to the nature of factor construction, correlation with the market factor is almost unavoidable when factors are returns. However, one should still be mindful of possible collinearity between SMB and market factors when estimating and testing. On an equally-weighted basis we see a very strong correlation of 0.72 between SMB and HML. This almost certainly will distort the relationship when used simultaneously within a regression. In particular it indicates that individual inferences regarding regression coefficients cannot be made with a high degree of certainty when estimating the FF3F model using equally-weighted factors.

The market and SMB factors are even more correlated when the sample period is shorted. The only reprieve available is to be mindful of possible collinearity between SMB and market factors when estimating and testing. On the other hand, coskewness and cokurtosis factors in general have very low correlations with each other and other factors allowing the study to better isolate the effects of each factor in the regression.

5.2 Distribution Analysis

The assumption of multivariate normality is central to the CAPM. OLS estimation also assumes that variables are normally distributed. Given the pervasive use of both the CAPM and OLS estimation in finance and economics, specifically asset pricing, one would expect asset returns to follow a normal distribution. Yet the literature reviewed in Section2.4 suggests otherwise. The primary purpose of this paper is to assess the validity of higher moment asset pricing models. It is therefore imperative to establish if higher moments exist. This section provides an in-depth analysis of both individual portfolio return distribution and joint return distributions.

Value-weighted Portfolio Analysis

Measures of location, dispersion and shape are calculated for the value weighted portfolios over the period January 1993 and December 2013 and are presented in Table 5.3. Recall **H_0**:

$$H_0: R_i \sim N(.)$$

Table 5.3 Return distribution analysis of value weighted portfolios (January 1993 - December 2013)

JSE listed share returns are sorted into sample portfolios for the January 1993 to December 2013. Sample portfolios are formed using either independent decile sorts or two-way quintile sorts on size (ME) and BE/ME. The result is 10 portfolios formed on size, 10 portfolios formed on value and 25 portfolios formed on size and value. Measures of location, dispersion and shape are calculated for sample portfolios and displayed in the table below. SMB and HML refer to the factors constructed within the FF3F model. J_{JB} refers to the Jarque-Bera test statistics. ***, **, * refer to statistical significance at the 1, 5 and 10% levels, respectively.

Factors	Mean (%)	Std. Dev.(%)	Skewness	Coskewness	Kurtosis	Cokurtosis	J_{JB}
Market Factor	0.707**	5.613	-0.764***		6.706***		156.752***
Risk-free rate	0.792***	0.264	0.445***	0.074***	2.63***	0.082	9.842***
SMB	0.742*	6.034	0.472***	-1.383***	5.17***	-1.244	53.819***
HML	0.731*	6.369	0.686***	1.091*	4.501***	-6.858	40.253***
Size Decile	**Mean (%)**	**Std. Dev.(%)**	**Skewness**	**Coskewness**	**Kurtosis**	**Cokurtosis**	**J_{JB}**
1	2.832***	8.341	0.912***	-0.782	7.775***	2.685	256.013***
2	1.896***	6.464	0.198	-1.122*	5.541***	-3.604	63.288***
3	1.859***	5.019	0.231	-0.644*	5.234***	-6.67***	49.592***
4	1.552***	5.35	-0.388***	-1.13***	5.562***	-4.437	68.935***
5	1.586***	4.655	-1.049***	-1.249***	9.821***	1.45	500.561***
6	1.397***	4.963	-0.904***	-0.822***	7.239***	0.305	207.931***
7	1.685***	5.097	-0.812***	-0.646*	8.138***	3.333	284.276***
8	1.656***	4.654	-0.838***	-0.606***	6.024***	-0.367	116.698***
9	1.436***	5.256	-0.457***	0.029	4.567***	-1.493	31.448***
10	1.419***	5.732	-0.466***	0.404***	5.07***	-1.228*	49.462***
BE / ME Decile	**Mean (%)**	**Std. Dev.(%)**	**Skewness**	**Coskewness**	**Kurtosis**	**Cokurtosis**	**J_{JB}**
1	1.144**	7.467	-0.599***	-0.457	6.741***	0.903	150.054***
2	1.491***	6.132	-0.411***	0.451	5.061***	-0.844	47.137***
3	1.358***	5.733	-0.967***	-0.73***	7.156***	-2	205.849***
4	1.552***	5.74	-0.245	0.369	3.833***	-3.066	8.564*
5	1.512***	6.136	-0.202	0.554*	4.388***	-0.655	19.49***
6	1.703***	6.638	0.363***	0.707*	6.265***	4.431	108.159***
7	1.546***	5.771	0.116	0.518	5.368***	1.717	53.973***
8	2.108***	6.277	-0.526***	-0.107	5.294***	-8.027***	61.417***
9	2.24***	6.92	0.388***	0.921*	3.579***	-7.828*	8.975*
10	2.005***	6.959	0.653***	0.896*	4.939***	-1.731	52.989***
Size-Value Quintiles	**Mean (%)**	**Std. Dev.(%)**	**Skewness**	**Coskewness**	**Kurtosis**	**Cokurtosis**	**J_{JB}**
S1V1	1.336***	7.505	-0.195	-1.494***	5.163***	0.272	45.95***
S1V2	2.093***	8.851	-0.033	-0.724	7.322***	-3.66	181.365***
S1V3	2.142***	8.2	0.595***	-1.915***	7.463***	-5.494	208.075***
S1V4	3.316***	11.12	1.493***	-0.557	7.921***	4.664	327.165***
S1V5	2.477***	11.85	3.762***	-0.671	30.009***	-11.327	7798.4***
S2V1	1.323***	7.092	-0.444***	-1.898***	5.658***	-3.609	75.693***

S2V2	1.333***	6.089	-0.126	-1.238***	5.559***	0.672	63.229***
S2V3	1.827***	5.657	0.011	-0.208	3.676***	-2.627	3.911
S2V4	2.275***	7.693	2.07***	-1.01	14.565***	-7.92	1492.693***
S2V5	2.009***	7.593	1.068***	0.187	6.978***	-9.302*	200.144***
S3V1	1.244***	5.739	-1.338***	-1.945***	11.835***	4.288	839.985***
S3V2	1.516***	5.795	-0.235	-0.812*	6.884***	2.708	148.316***
S3V3	1.233***	5.676	-0.661***	-0.887***	6.863***	1.532	162.427***
S3V4	1.832***	5.623	-0.297**	-0.771*	4.579***	-2.631	26.872***
S3V5	1.578***	5.711	-0.277**	-0.218	4.492***	-4.85	23.826***
S4V1	1.287***	6.127	-0.972***	-1.26***	8.902***	2.131	378.965***
S4V2	1.386***	5.382	-0.909***	-1.355***	6.794***	-1.494	173.13***
S4V3	1.504***	4.765	-0.267**	-0.149	4.924***	3.549	37.827***
S4V4	1.732***	5.57	-0.544***	-0.8*	7.244***	1.408	186.904***
S4V5	2.381***	5.838	0.581***	0.994***	4.051***	-4.213	23.805***
S5V1	1.066**	7.157	-0.439***	0.022	6.16***	2.298	104.061***
S5V2	1.412***	6.025	-0.554***	-0.142	5.498***	-4.479*	72.175***
S5V3	1.472***	5.655	-0.066	0.868***	3.842***	0.18	6.437*
S5V4	1.616***	6.515	-0.018	0.681***	4.981***	1.946	37.105***
S5V5	1.621***	6.301	-0.099	0.589*	5.348***	-2.814	52.91***

The sample consists of 45 value weighted portfolios over the January 1993 to December 2013 period. Skewness is significantly different from that of a normal distribution for 32 of the 45 sample portfolios while kurtosis is statistically significant in all 45 cases. The final column of Table 5.3 shows the results of the Jarque-Bera test statistic, a formal test of univariate normality. We reject $\mathbf{H_0}$ at the 99 percent confidence level for all ten size deciles, eight value deciles and 23 of the 25 size-value sorted portfolios. Furthermore we reject $\mathbf{H_0}$ for the two remaining value deciles and one of the size-value sorted portfolios at the 90 percent confidence level. Collectively the hypothesis of normality is rejected for all but one of the value-weighted sample portfolios for the period January 1993 and December 2013.

Measures of location, dispersion and shape are calculated for the value weighted portfolios over the period January 1997 and December 2013 and are presented in Table 5.4. The sample is extended by the inclusion of industry portfolios as well conditional systematic skewness and kurtosis portfolios. The sample consists of 34 value weighted portfolios over the January 1997 and December 2013 period. Skewness is significantly different from that of a normal distribution for 28 of the 34 sample portfolios while kurtosis is statistically significant in all cases. Using the Jarque-Bera test statistic we reject $\mathbf{H_0}$ at the 99 percent confidence level for all 34 value weighted portfolios. Thus we reject the hypothesis of normality for all value-weighted sample portfolios for the period January 1997 to December 2013.

The pervasiveness of leptokurtic return distributions within our sample data confirms that JSE listed firms are more prone to extreme events than would be predicted by a normal distribution. The pervasiveness of skewness is more interesting as it does not affect all sample portfolios in a uniform manner as with kurtosis. The high return portfolios, small size and high value, tend to be positively skewed while low return portfolios, big size and growth (low value), tend to be negatively skewed.

Table 5.4 Return distribution analysis of value weighted portfolios (January 1997 - December 2013)

JSE listed share returns are sorted into sample portfolio for the January 1997 to December 2013. Sample portfolios are formed using either independent decile sorts on size (ME) and BE/ME. The result is 10 portfolios formed on size and 10 portfolios formed on value. Six portfolios are formed using univariate sorts by ranking firms based on coskewness and cokurtosis. Finally industry portfolios are added to the sample. Measures of location, dispersion and shape are calculated for sample portfolios and displayed in the table below. SMB and HML refer to the factors constructed within the FF3F model while S^- (K^+) refer proxy portfolio for conditional systematic skewness (kurtosis). J_{JB} refers to the Jarque-Bera test statistic. ***, **, * refer to statistical significance at the 1, 5 and 10% levels, respectively.

Factors	Mean (%)	Std. Dev.(%)	Skewness	Coskewness	Kurtosis	Cokurtosis	J_{JB}
Market	0.683*	5.828	-0.857***		6.517***		119.042***
SMB	0.082	5.207	-0.21	-1.663***	3.401***	-0.85	2.376
HML	0.347	6.308	0.639***	0.025*	4.674***	0.244	34.122***
S^-	0.773*	6.163	-0.775***	0.123	5.986***	0.994	87.748***
K^+	1.345***	6.534	-0.445***	0.015*	4.681***	0.454	27.362***
Size Decile	**Mean (%)**	**Std. Dev.(%)**	**Skewness**	**Coskewness**	**Kurtosis**	**Cokurtosis**	**J_{JB}**
1	1.859***	6.715	0.423	-1.367***	4.336***	0.888	12.955***
2	1.256***	5.487	-0.651***	-1.254***	5.141***	0.383	48.324***
3	1.434***	4.466	-0.489***	0.023*	3.812***	0.011*	12.38***
4	1.192***	5.09	-0.812***	-1.2***	5.95***	0.129	88.018***
5	1.417***	4.73	-1.264***	-1.335***	10.549***	0.709	497.651***
6	0.982***	4.877	-1.21***	-0.917***	8.347***	0.947	270.064***
7	1.491***	5.05	-1.033***	-0.829***	9.549***	0.08	369.209***
8	1.578***	4.701	-0.931***	-0.696***	6.36***	0.835	114.979***
9	1.357***	5.399	-0.482***	0.404	4.482***	0.6	23.714***
10	1.372***	5.918	-0.557***	0.425***	4.675***	-1.847***	30.954***
BE / ME Decile	**Mean (%)**	**Std. Dev.(%)**	**Skewness**	**Coskewness**	**Kurtosis**	**Cokurtosis**	**J_{JB}**
1	1.097*	7.975	-0.567***	0.414	6.197***	0.169	88.668***
2	1.448***	6.461	-0.425***	0.062	4.775***	0.991	29.268***
3	1.366***	6.059	-1.015***	0.011*	6.92***	0.215	152.17***
4	1.542***	5.914	0.044*	0.088	3.479***	0.035*	4.19
5	1.455***	6.399	0.135	0.783***	4.241***	0.239	12.218***
6	1.463***	6.497	0.321	0.194	5.771***	0.352	58.481***
7	1.445***	5.871	0.363	0.089	5.653***	0.534	53.443***
8	1.766***	6.166	-0.719***	0.472	5.787***	-8.723***	76.031***
9	1.852***	6.574	0.089	1.479***	3.239***	0.821	1.989
10	1.543***	6.276	0.552***	0.088	5.785***	0.578	68.943***
Industry Portfolios	**Mean (%)**	**Std. Dev.(%)**	**Skewness**	**Coskewness**	**Kurtosis**	**Cokurtosis**	**J_{JB}**
J500	1.399**	8.006	0.017*	1.282***	5.998***	0.182	72.747***
J510	1.449***	6.134	-0.683***	0.138	5.418***	0.859	59.507***
J520	1.898***	6.779	0.078	0.407	4.515***	0.472	18.718***
J530	1.679***	6.169	-0.5***	0.127	4.528***	0.943	25.322***
J540	1.616***	6.595	-0.917***	0.036*	6.161***	0.572	103.988***
J550	2.056***	9.32	0.035*	0.293	5.312***	0.454	43.369***
J560	1.305***	6.319	-1.04***	-1.232***	11.966***	0.012*	664.837***

J570	0.977	10.64	-0.523***	0.475	5.205***	0.686	45.519***
Systematic Skewness & Kurtosis Portfolios	Mean (%)	Std. Dev.(%)	Skewness	Coskewness	Kurtosis	Cokurtosis	J_{JB}
S^-	0.773*	6.163	-0.775***	0.123	5.986***	0.994	87.748***
M	1.286***	5.75	-0.487***	0.086	4.712***	0.762	29.476***
S^+	1.743***	6.895	-0.568***	0.145	5.072***	0.124	42.765***
				0.015			
K^-	1.592***	6.441	-0.583***	0.174	5.643***	0.636	64.106***
M	1.392***	5.742	-0.823***	0.335	5.981***	0.37	90.06***
K^+	1.345***	6.534	-0.445***	0.015*	4.681***	0.454	27.362***

However, the nature of the causal relationship remains unclear. For example, can an investor consistently obtain higher returns by investing in positively skewed shares? Or are positively skewed return distributions simply reflecting the history of impressive firm results. The methodology of the study does not facilitate a clear answer as results are presented in an ex-post manner.

Tables 5.3 and 5.4 also show unconditional coskewness and cokurtosis as developed by Kraus and Litzenberger (1976) and Fang and Lai (1997) for sample portfolios. Unconditional coskewness is statically significant for eight of the ten size sorted portfolio and 15 out of the 25 size-value sorted portfolios but only four of the ten value portfolios over the sample period January 1993 to December 2013. This suggests that market equity and unconditional coskewness are related in some manner but there is no clear indication as to the nature of the relationship. Unconditional coskewness is statically significant for nine of the ten size sorted portfolios but only seven of the remaining 24 portfolios over the period January 1997 to December 2013. This again points to a possible relationship between coskewness and firm size. Barone-Adesi, Gagliardini and Urga (2004) find that firm size and coskewness are correlated and suggest that the anomalous relationship between size and expected return may be explained by the omission of coskewness from the CAPM.

Unconditional cokurtosis is statistically significant for only six out of 45 sample portfolios over the January 1993 to December 2013 sample period and for only five of the 34 portfolios over January 1997 to December 2013 sample period. These results, in combination with the results of the factors analysis, suggest cokurtosis with the market portfolio has little if any explanatory power related to the time-series of JSE listed stock returns.

Table 5.3 and 5.4 also enable a comparison of the size and value premia when a decile sorting methodology is used rather than a quintile sorting methodology. The mean return on the smallest size decile is the highest by some margin within the value weighted portfolio set over the sample period January 1993 to December 2013. The sorting of portfolios into deciles results in a “micro” portfolio where very small shares are concentrated. The size premium when measured as the arbitrage portfolio between a long position in the micro portfolio and short position in the largest size decile

equals 1.41 percent. The decile equivalent to the value effect is only slightly larger than its quintile counterpart. We document the decile value effect at 0.86 percent per month.

Sorting into decile rather than quintile portfolios appears to amplify the magnitude of the size and value premia. The micro portfolio, even though it is subject to a liquidity filter, would contain shares that are traded the least within the sample. Standard deviation is highest in the smallest size portfolio decile. This could be seen to provide further support to the illiquidity hypothesis as shares that are not traded regularly are more susceptible to market demand. The amplification of the decile equivalent of the value premium can be explained in light of Fama and French's (1996) hypothesis of financial distress. The highest value decile stocks will have higher BE / ME ratios than the highest value quintile stocks as a simple consequence of sorting. The higher a firm's BE / ME ratio the higher its level of financial distress. As per Fama and French's (1996) hypothesis the higher the level of firm financial distress the higher the return should the firm survive the financial distress.

Recalling that the size effect all but disappears for quintile sorts over the January 1997 to December 2013 period, it is interesting that the "micro effect" is 0.49 percent per month. Increasing the number of portfolios formed out of the universe of shares thus seems to increase the magnitude of the size premium. This highlights the importance of methodology when completing portfolio sorts.

Finally it should be noted that the portfolio set is extended by the inclusion of conditional systematic skewness and kurtosis portfolios for the January 1997 to December 2013 period. Harvey and Siddique (2000) find that the regression coefficient for coskewness is significant when using the excess return on S^- rather than the spread between S^- and S^+ as a proxy. We will therefore use the excess return on S^- as our proxy for conditional coskewness when estimating the conditional quadratic CAPM. Similarly we will use the excess return on K^+ as our proxy for conditional cokurtosis when estimating the cubic CAPM.

Collectively the results of the value weighted portfolio distribution analysis provides very strong evidence that JSE share returns are not normally distributed, or at least when they are sorted into some form of portfolio. Furthermore we saw reasonable support for the unconditional quadratic CAPM and that coskewness is priced.

Equally-weighted Portfolio Analysis

The distribution analysis of the equally weighted portfolios produces results very similar to that of the value weighted portfolios. Measures of location, dispersion and shape for the sample equally weighted portfolios for period January 1993 to December 2013 are presented in Table 5.5. The equally weighted set consist of 20 portfolios. Skewness is significantly different from that of a normal distribution for 17 of the 20 sample portfolios. All sample portfolios exhibit kurtosis statistically different from that of a normal distribution. Using the Jarque-Bera test statistic we reject $\mathbf{H_0}$ at the 99 percent confidence level for all 20 equally weighted portfolios for the period January 1993 to December 2013.

Table 5.5 Return distribution analysis of equally weighted portfolios (January 1993 - December 2013)

JSE listed share returns are sorted into sample portfolios for the January 1993 to December 2013. Sample portfolios are formed using either independent decile sorts on size (ME) and BE/ME. The result is 10 portfolios formed on size and 10 portfolios formed on value. Measures of location, dispersion and shape are calculated for sample portfolios and displayed in the table below. SMB and HML refer to the factors constructed within the FF3F model. J_{JB} refers to the Jarque-Bera test statistic. ***, **, * refer to statistical significance at the 1, 5 and 10% levels, respectively.

Factors	Mean (%)	Std. Dev.(%)	Skewness	Coskewness	Kurtosis	Cokurtosis	J_{JB}
SMB	0.999**	6.342	1.257***	-1.09*	7.791***	0.766	288.095***
HML	1.457***	6.137	1.744***	0.946*	10.603***	-2.979	691.417***
Size Decile	**Mean (%)**	**Std. Dev.(%)**	**Skewness**	**Coskewness**	**Kurtosis**	**Cokurtosis**	**J_{JB}**
1	2.773***	8.926	1.29***	-0.478	7.72***	4.076	284.885***
2	1.96***	7.139	1.123***	-1.129*	9.944***	-5.892	523.947***
3	1.763***	5.237	0.106	-0.755*	5.156***	-5.766*	44.555***
4	1.492***	5.619	0.006	-1.038***	5.813***	-4.579	75.877***
5	1.45***	4.521	-0.913***	-1.063***	9.123***	1.678	400.513***
6	1.628***	4.881	-0.405***	-0.575*	5.484***	1.937	65.621***
7	1.669***	5.099	-0.564***	-0.465	6.548***	2.115	134.638***
8	1.7***	4.685	-0.587***	-0.352	4.714***	-1.3	41.64***
9	1.471***	5.195	-0.336**	0.253	3.835***	-1.958	10.771***
10	1.319***	5.311	-0.408***	0.399***	4.881***	-0.984	40.188***
BE / ME Decile	**Mean (%)**	**Std. Dev.(%)**	**Skewness**	**Coskewness**	**Kurtosis**	**Cokurtosis**	**J_{JB}**
1	1.311***	5.706	-0.954***	-1.333***	9.186***	0.961	411.44***
2	1.266***	5.099	-0.711***	-0.608*	5.649***	-0.637	87.905***
3	1.311***	4.807	-0.821***	-0.684***	7.213***	1.171	199.86***
4	1.322***	4.918	-0.564***	-0.363	5.14***	-0.296	56.498***
5	1.447***	4.797	-0.677***	-0.625*	6.203***	0.092	117.641***
6	1.479***	5.304	-0.281**	-0.214	4.843***	0.25	35.226***
7	1.81***	5.085	-0.055	0.141	4.041***	-4.585*	9.897***
8	2.044***	5.61	-0.436***	-0.66*	5.175***	-4.877*	52.714***
9	2.713***	7.349	1.148***	-0.99*	10.239***	-2.646	567.421***
10	2.62***	8.623	2.083***	0.307	11.651***	-2.853	912.691***

Measures of location, dispersion and shape for sample equally weighted portfolios for the period January 1997 to December 2013 are presented in Table 5.6. The equally-weighted portfolios' skewness and kurtosis are more prevalent compared to value-weighted portfolios as higher moments are significantly different from that of a normal distribution in all cases. Using the Jarque-Bera test statistic we reject $\mathbf{H_0}$ at the 99 percent confidence level for all but two of the equally weighted portfolios for the period January 1997 to December 2013. For both sample periods unconditional coskewness is statically significant for 12 of the 20 portfolios while cokurtosis is significant only three times. As with the value-weighted portfolios extreme events are more common than would be predicted by a normal distribution and high (low) return portfolios tend to be positively (negatively) skewed.

Table 5.6 Return distribution analysis of equally weighted portfolios (January 1997 - December 2013)

JSE listed share returns are sorted into sample portfolios for the January 1997 to December 2013. Sample portfolios are formed using either independent decile sorts on size (ME) and BE/ME. The result is 10 portfolios formed on size and 10 portfolios formed on value. Finally six portfolios are formed using univariate sorts by ranking firms based on coskewness and cokurtosis. Measures of location, dispersion and shape are calculated for sample portfolios and displayed in the table below. SMB and HML refer to the factors constructed within the FF3F model while SKEW (KURT) refer proxy portfolio for systematic skewness (kurtosis). J_{JB} refers to the Jarque-Bera test statistic. ***, **, * refer to statistical significance at the 1, 5 and 10% levels, respectively.

Factors	Mean (%)	Std. Dev.(%)	Skewness	Coskewness	Kurtosis	Cokurtosis	J_{JB}
SMB	0.14	4.921	0.344**	-1.651***	4.49***	-1.235	20.113***
HML	0.707**	4.309	0.031*	0.062	4.864***	0.381	29.111***
S^-	1.59***	5.124	-1.083***	-0.938***	7.559***	0.593	199.244***
K^+	1.777***	5.161	-0.833***	-0.715***	7.061***	0.872	149.912***
Size Decile	**Mean (%)**	**Std. Dev.(%)**	**Skewness**	**Coskewness**	**Kurtosis**	**Cokurtosis**	**J_{JB}**
1	1.774***	7.209	0.993***	0.014*	9.035***	0.841	315.697***
2	1.124***	5.307	-0.546***	-1.218***	5.303***	0.232	49.755***
3	1.238***	4.698	-0.644***	-0.826***	4.19***	0.027*	23.804***
4	1.079***	5.142	-0.637***	-1.081***	5.389***	0.262	56.421***
5	1.23***	4.535	-1.25***	-1.263***	9.955***	0.999	428.717***
6	1.272***	4.726	-0.645***	0.016*	6.331***	0.364	98.598***
7	1.422***	4.962	-0.826***	0.02*	7.683***	0.262	192.062***
8	1.579***	4.627	-0.741***	0.046*	4.961***	0.418	46.785***
9	1.377***	5.274	0.017*	0.131	3.74***	0.537	7.967*
10	1.288***	5.478	-0.472***	0.439***	4.464***	0.255	22.975***
BE / ME Decile	**Mean (%)**	**Std. Dev.(%)**	**Skewness**	**Coskewness**	**Kurtosis**	**Cokurtosis**	**J_{JB}**
1	1.024**	5.859	-1.023***	-1.306***	9.505***	0.445	364.117***
2	1.1***	5.274	-0.785***	0.02*	5.679***	0.657	74.78***
3	1.164***	5	-0.897***	-0.771***	7.222***	0.941	163.969***
4	1.082***	5.084	-0.618***	0.094	4.997***	0.584	42.368***
5	1.27***	4.678	-0.937***	-0.752***	7.225***	0.884	166.581***
6	1.186***	4.909	-0.66***	0.089	5.234***	0.974	51.874***
7	1.611***	4.672	0.04*	0.43	3.634***	0.015*	5.521
8	1.706***	5.285	-0.781***	0.012*	6.001***	0.022*	88.723***
9	1.895***	6.294	1.161***	0.029*	15.606***	0.814	1293.297***
10	1.617***	5.8	0.942***	0.291	7.884***	0.338	213.907***
Coskewness & Cokurtosis	**Mean (%)**	**Std. Dev.(%)**	**Skewness**	**Coskewness**	**Kurtosis**	**Cokurtosis**	**J_{JB}**
S-	1.59***	5.124	-1.083***	-0.938***	7.559***	0.593	199.244***
M	1.769***	5.115	-1.089***	-0.902***	8.11***	0.743	241.459***
S+	1.963***	5.272	-0.791***	0.03*	6.169***	0.969	97.299***
K^-	1.912***	5.191	-0.986***	-0.975***	7.679***	0.548	201.342***
M	1.563***	5.064	-1.09***	-0.72***	7.101***	0.366	168.773***
K^+	1.777***	5.161	-0.833***	-0.715***	7.061***	0.872	149.912***

As with the value weighted portfolios, sorting into deciles rather than quintiles exaggerates the magnitude of the size effect. The micro effect is 1.45 percent per month compared to the 1.00 percent quintile equivalent. Unlike the value weighted portfolios the decile equivalent of the value effect however is slightly smaller at 1.31 percent per month. Sorting into deciles only appears to increase the magnitude of the size effect. This is confirmed in the January 1997 to December 2013 sample period where the micro effect is 0.49 percent per month while the decile equivalent of the value effect is 0.53 percent per month.

The results of the distribution analysis of the equally weighted portfolios strongly indicate that JSE listed share returns are not normally distributed. Furthermore the results indicate that unconditional coskewness is important in pricing equities.

Table 5.7 Multivariate tests of normality

JSE listed share returns are sorted into sample portfolios for the January 1997 to December 2013. Sample portfolios are formed using either independent decile sorts or two-way quintile sorts on size (ME) and BE/ME. The result is 10 portfolios formed on size, 10 portfolios formed on value and 25 portfolios formed on size and value. Six portfolios are formed using univariate sorts by ranking firms based on coskewness and cokurtosis. Finally industry portfolios are added to the sample. Table 5.7 displays the test statistic of multivariate normality described Section4.2 for sample portfolio panels. ***, **, * corresponds to a rejection of the hypothesis that a given panel is consistent with a joint normal distribution at the 0.01, 0.05 or 0.1 significance level respectively.

Value Weighted Panels	Shapiro-Wilk	Henze-Zirkler	Mardia's Skewness	Adjusted Mardia's Skewness	Mardia's Kurtosis	Doornik-Hansen
A) Size	0.9***	1.245***	15.903***	15.903***	165.228***	209.417***
B) Value	0.831***	1.461***	19.831***	19.831***	175.195***	234.721***
C) Systematic Skewness	0.964***	1.648***	1.04***	1.04***	21.211***	29.787***
D) Systematic Kurtosis	0.959***	1.595***	1.546***	1.546***	21.248***	27.886***
E) Size-Value	0.724***	1.01***	186.586***	186.586***	903.055***	1036.385***
F) Industry	0.917***	1.591***	13.39***	13.39***	121.131***	222.91***

Equally Weighted Panels	Shapiro-Wilk	Henze-Zirkler	Mardia's Skewness	Adjusted Mardia's Skewness	Mardia's Kurtosis	Doornik-Hansen
A) Size	0.863***	1.314***	24.93***	24.93***	181.969***	323.104***
B) Value	0.786***	1.249***	28.772***	28.772***	176.017***	443.496***
C) Systematic Skewness	0.937***	1.088**	2.551***	2.551***	25.886***	77.617***
D) Systematic Kurtosis	0.943***	1.367***	2.434***	2.434***	25.299***	63.754***

Multivariate Normality

The assumption of normality underlying the CAPM does not correspond to a single asset but rather that assets are jointly normally distributed. Table 5.7 therefore illustrates the six multivariate normality tests discussed in Section4.2 applied to sample portfolios within their respective panels. The results provide very strong evidence that sample portfolio returns are not jointly normally distributed. We reject **H_0** with 99 percent confidence for all panels using six different test statistics.

The results of the univariate and multivariate normality tests strongly indicate that JSE listed equity returns do not follow a normal distribution or at the very least when sorted into portfolios are not normally distributed. The estimation and testing methodology when evaluating asset pricing models on the JSE should therefore account for non-normality by design. Furthermore the result invalidate one of the central assumptions of the CAPM. Strictly speaking the CAPM therefore does not hold for JSE data.

5.3 Time-Series Regressions

Sample portfolios are regressed over two periods as conditional asset pricing models require 60 months lead time to compile factors and data for FTSE / JSE industry indices are only available from January 1997. The two sample periods therefore are: (1) January 1993 to December 2013; and (2) January 1997 to December 2013. The size sorted, value sorted, size-value sorted and industry portfolio returns are individually regressed on respective asset pricing model factors:

(A) CAPM:

$$R_t^{ei} = a^i + \beta_i^m\left(R_t^m - R_t^f\right) + \varepsilon_t^i \qquad t = 1, \dots, T \qquad (5.1)$$

Where R_t^{ei} is the excess return of portfolio i over the risk-free rate (R^f); a^i is the regression intercept; R_t^m is the return on the market portfolio; β_i^m is the regression coefficient on the market factor for asset i; and ε_t^i is the regression residual.

(B) FF3F model:

$$R_t^{ei} = a^i + \beta_i^m\left(R_t^m - R_t^f\right) + \beta_i^s SMB_t + \beta_i^v HML_t + \varepsilon_t^i \qquad t = 1, \dots, T \qquad (5.2)$$

Where SMB_t is the return on the proxy portfolio for the size factor and β_i^s is the regression coefficient on the size factor; HML_t is the return on the proxy portfolio for the value factor and β_i^v is the regression coefficient on the size factor

(C) Unconditional quadratic CAPM (CAPM^2):

$$R_t^{ei} = a^i + \beta_i^m\left(R_t^m - R_t^f\right) + \beta_i^{m2}\left(R_t^m - R_t^f\right)^2 + \varepsilon_t^i \qquad t = 1, \dots, T \qquad (5.3)$$

Where β_i^{m2} is the regression coefficient on the market factor squared for asset i.

(D) Unconditional cubic CAPM (CAPM^3):

$$R_t^{ei} = a^i + \beta_i^m\left(R_t^m - R_t^f\right) + \beta_i^{m2}\left(R_t^m - R_t^f\right)^2 + \beta_i^{m3}\left(R_t^m - R_t^f\right)^3 + \varepsilon_t^i \qquad t = 1, \dots, T \quad (5.4)$$

Where β_i^{m3} is the regression coefficient on the market factor cubed for asset i.

(E) Conditional quadratic CAPM (CCAPM^2):

$$R_t^{ei} = a^i + \beta_i^m\left(R_t^m - R_t^f\right) + \beta_i^{sks}\left(S_t^- - R_t^f\right) + \varepsilon_t^i \qquad t = 1, \dots, T \qquad (5.5)$$

Where $\left(S_t^- - R_t^f\right)$ is the return on the proxy portfolio for the conditional coskewness factor and β_i^{sks} is its regression coefficient for asset i.

(F) Conditional quadratic CAPM (CCAPM^3):

$$R_t^{ei} = a^i + \beta_i^m\left(R_t^m - R_t^f\right) + \beta_i^{sks}\left(S_t^- - R_t^f\right) + \beta_i^k\left(K_t^+ - R_t^f\right) + \varepsilon_t^i \qquad t = 1, \dots, T \quad (5.6)$$

Where $\left(K_t^+ - R_t^f\right)$ is the return on the proxy portfolio for the conditional cokurtosis factor and β_i^k is its regression coefficient for asset i. Furthermore APT models are estimated by combining FF3F and higher moment CAPM factors. This produces a very large set of regression results of which only the most poignant are represented in this section. The complete set of cross-sectional regression results is presented. This further reduces the need to share all the time-series regression results in order to avoid redundancies.

Time-series Variation

The R-squared statistic of a time-series regression relates how well independent variables explain variation in the dependent variable over time. This study is assessing the ability of factor proxy portfolio returns to explain the variation in sample portfolio returns. Fama and French (1993) found that the addition of the SMB and HML proxy portfolios greatly improved the ability of the CAPM to explain variation in returns. Harvey and Siddique (2000) and Dittmar (2002) found similar results by adding higher co-moments to the CAPM instead. The 25 size-value sorted portfolio set is used frequently when assessing the ability of asset pricing to explain variation in returns.

Table 5.8 displays the adjusted R-squared for each of the 25 size-value portfolios when estimating the various asset pricing models for the shorter sample period. The CAPM performs poorly in that only 11 out of the 25 portfolios have an R-squared of above 0.4 and only five above 0.5. The CAPM performs best in the high value quintile producing the highest R-squared value 0.745. The FF3F model does remarkably better producing R-squared values of above 0.4 for all but three of the 25 portfolios while 13 of the 25 portfolio have R-squared values above 0.5. The effect is most pronounced in the lower value quintiles. This provides support for the hypothesis that the size premium is disappearing as evidenced in the results of the portfolios sorts. Furthermore the R-squared values for both the CAPM and FF3F model are substantially lower than the Fama and French (1993) results. This suggests that the models have lost explanatory power due to either the differing sample periods or that the models are less appropriate for JSE return data.

The unconditional quadratic CAPM performs marginally better than the CAPM producing R-squared values of above 0.4 for 12 out of the 25 portfolios and seven above 0.5. The highest value coincidentally being 0.745 again. The unconditional cubic CAPM produces results that, in most cases, are identical to that of the unconditional quadratic CAPM to the second decimal suggesting that the

addition of a cubic term does not make a large difference in explaining variation in returns over time. This is in line with the previous set of results where the unconditional cokurtosis was statistically insignificant in most cases. This leads the study to conjecture that even though kurtosis is highly present in asset returns, cokurtosis has little explanatory power not already captured by coskewness.

Table 5.8 Adjusted R-squared statistics of time-series regressions

JSE listed share returns are sorted into sample portfolio for the period January 1997 to December 2013. The portfolios are formed using two-way quintile sorts on size (ME) and BE/ME, respectively. Sample portfolio returns are regressed on the respective factor portfolios ($\forall t$) for the following asset pricing models:

(A) CAPM $R_t^{ei} = a^i + \beta_i^m\left(R_t^m - R_t^f\right) + \varepsilon_t^i$

(B) FF3F model $R_t^{ei} = a^i + \beta_i^m\left(R_t^m - R_t^f\right) + \beta_i^s SMB_t + \beta_i^v HML_t + \varepsilon_t^i$

(C) unconditional quadratic CAPM (CAPM^2) $R_t^{ei} = a^i + \beta_i^m\left(R_t^m - R_t^f\right) + \beta_i^{m2}\left(R_t^m - R_t^f\right)^2 + \varepsilon_t^i$

(D) unconditional cubic CAPM (CAPM^3) $R_t^{ei} = a^i + \beta_i^m\left(R_t^m - R_t^f\right) + \beta_i^{m2}\left(R_t^m - R_t^f\right)^2 + \beta_i^{m3}\left(R_t^m - R_t^f\right)^3 + \varepsilon_t^i$

(E) conditional quadratic CAPM (CCAPM^2) $R_t^{ei} = a^i + \beta_i^m\left(R_t^m - R_t^f\right) + \beta_i^{sks}\left(S_t^- - R_t^f\right) + \varepsilon_t^i$

(F) conditional quadratic CAPM (CCAPM^3) $R_t^{ei} = a^i + \beta_i^m\left(R_t^m - R_t^f\right) + \beta_i^{sks}\left(S_t^- - R_t^f\right) + \beta_i^k\left(K_t^+ - R_t^f\right) + \varepsilon_t^i$

The table presents the adjusted R-squared statistics of the various regressions.

CAPM

	BE/ME quintiles				
Size quintiles	Low	2	3	4	High
Small	0.204	0.360	0.337	0.446	0.695
2	0.157	0.264	0.371	0.478	0.705
3	0.236	0.275	0.481	0.482	0.701
4	0.131	0.203	0.376	0.487	0.745
Big	0.065	0.227	0.392	0.441	0.712

FF3F

	BE/ME quintiles				
Size quintiles	Low	2	3	4	High
Small	0.475	0.446	0.468	0.531	0.832
2	0.490	0.430	0.450	0.595	0.742
3	0.471	0.441	0.574	0.552	0.728
4	0.416	0.370	0.521	0.560	0.778
Big	0.281	0.347	0.517	0.622	0.785

CAPM^2

	BE/ME quintiles				
Size quintiles	Low	2	3	4	High
Small	0.223	0.399	0.397	0.471	0.693
2	0.169	0.298	0.382	0.514	0.704
3	0.274	0.277	0.495	0.483	0.711
4	0.133	0.213	0.400	0.509	0.745
Big	0.071	0.223	0.390	0.446	0.716

CAPM^3

	BE/ME quintiles				
Size quintiles	Low	2	3	4	High
Small	0.220	0.405	0.398	0.471	0.693
2	0.172	0.294	0.383	0.512	0.707
3	0.274	0.278	0.496	0.482	0.710
4	0.139	0.212	0.408	0.510	0.744
Big	0.076	0.231	0.399	0.459	0.716

CCAPM^2

	BE/ME quintiles				
Size quintiles	Low	2	3	4	High
Small	0.209	0.381	0.368	0.588	0.704
2	0.175	0.324	0.394	0.543	0.716
3	0.243	0.289	0.495	0.529	0.706
4	0.129	0.250	0.403	0.521	0.744
Big	0.066	0.230	0.401	0.456	0.712

CCAPM^3

	BE/ME quintiles				
Size quintiles	Low	2	3	4	High
Small	0.206	0.379	0.381	0.590	0.717
2	0.171	0.322	0.393	0.547	0.727
3	0.240	0.285	0.501	0.528	0.705
4	0.125	0.257	0.403	0.523	0.744
Big	0.065	0.235	0.398	0.453	0.710

The conditional quadratic CAPM uses the excess return over the risk-free rate on the S^- portfolio rather than the spread between due S^- and S^+ as Harvey and Siddique (2000) found it to be significant

more often. Similarly the conditional cubic CAPM uses the excess return over the risk-free rate on the K^+ portfolio rather than the spread between K^+ and K^-. The conditional quadratic CAPM produces R-squared values are slightly better than that of the CAPM. Only 13 of the 25 portfolio have R-squared values above 0.4 and only nine above 0.5. These slightly better results suggest that the impact of coskewness with the market portfolio differs through time and conditional coskewness can better explain the time variation in returns compared to its unconditional counterpart. The difference in R-squared value is very small so we draw this inference with very little certainty. As with the unconditional version, the conditional cubic CAPM explains very little variation in portfolio returns over time not already captured by the conditional quadratic CAPM.

Cokurtosis is rarely significant and at best has a marginal impact. This is surprising as there are two major financial crises contained within the sample, the dotcom bubble and the subprime crises of 2008. Furthermore portfolios are all leptokurtic yet time-series variation appears to be captured by portfolio covariance and coskewness with the market portfolio rather than its cokurtosis. This leads the study to conjecture that the inclusion of the fourth term in a Taylor series expansion within the marginal rate of substitution is unnecessary.

5.4 Cross-sectional Regressions

The two crucial aspects determining the validity of an asset pricing model are whether factor risk premia are priced and whether pricing errors are zero. We address these aspects in the cross-sectional regressions in this Sectionover the two sample periods. The size sorted, value sorted, size-value sorted and industry portfolio returns are individually regressed on respective asset pricing model factors. The time-series regression estimates are summarised by Equation 5.7:

$$R_t^{ei} = a^i + \beta_i f_t' + \varepsilon_t^i \qquad t = 1, \dots, T \qquad (5.7)$$

We estimate the respective asset pricing models by regressing the expected returns of sample portfolios on the coefficients of model factors obtained from the time-series regressions:

(A) CAPM:

$$E_T(R^{ei}) = \beta_i^m \lambda^m + \alpha^i \qquad i = 1, \dots, N \qquad (5.8)$$

Where λ^m is the market risk premium calculated as the cross-sectional regression coefficient of market beta; and α^i is cross-sectional regression pricing error for asset i.

(B) Unconditional quadratic CAPM (CAPM^2):

$$E_T\left(R^{ei}\right) = \beta_i^m \lambda^m + \beta_i^{m2} \lambda^{m2} + \alpha^i \qquad i = 1, \dots, N \qquad (5.9)$$

Where λ^{m2} is the risk premium on coskewness or the cross-sectional regression coefficient of β_i^{m2}.

(C) Unconditional cubic CAPM (CAPM^3):

$$E_T\left(R^{ei}\right) = \beta_i^m \lambda^m + \beta_i^{m2} \lambda^{m2} + \beta_i^{m3} \lambda^{m3} + \alpha^i \qquad i = 1, \dots, N \qquad (5.10)$$

Where λ^{m3} is the risk premium on cokurtosis or the cross-sectional regression coefficient of β_i^{m3}.

(D) FF3F model:

$$E_T(R^{ei}) = \beta_i^m \lambda^m + \beta_i^s \lambda^s + \beta_i^v \lambda^v + \alpha^i \qquad i = 1, \dots, N \qquad (5.11)$$

Where λ^s and λ^v are the risk premia on the FF3F model size and value factors calculated as the cross-sectional regression coefficients of SMB and HML portfolios.

(E) APT

$$E_T(R^{ei}) = \beta_i^m \lambda^m + \beta_i^s \lambda^{vs} + \beta_i^v \lambda^v + \beta_i^{m2} \lambda^{m2} + \beta_i^{m3} \lambda^{m3} + \alpha^i \quad i = 1, \dots, N \qquad (5.12)$$

(F) Conditional quadratic CAPM (CCAPM^2)

$$E_T(R^{ei}) = \beta_i^m \lambda^m + \beta_i^{sks} \lambda^{sks} + \alpha^i \qquad i = 1, \dots, N \qquad (5.13)$$

Where λ^{sks} is the risk premium on conditional coskewness calculated as the cross-sectional regression coefficients of the S^- portfolio.

(G) Conditional cubic CAPM (CCAPM^3)

$$E_T(R^{ei}) = \beta_i^m \lambda^m + \beta_i^{sks} \lambda^{sks} + \beta_i^k \lambda^k + \alpha^i \qquad i = 1, \dots, N \qquad (5.14)$$

Where λ^{sks} is the risk premium on conditional cokurtosis calculated as the cross-sectional regression coefficients of the K^+ portfolio.

(H) APT including conditional factors (CAPT)

$$E_T(R^{ei}) = \beta_i^m \lambda^m + \beta_i^s \lambda^v + \beta_i^v \lambda^v + \beta_i^{sks} \lambda^{sks} + \beta_i^k \lambda^k + \alpha^i \qquad i = 1, \dots, N \qquad (5.15)$$

Cross-sectional Variation

The cross-sectional R-squared of an asset pricing model can be thought of as the models ability to explain variation in the expected returns of the asset being priced. The CAPM typically produces quite low cross-sectional R-squared values when expected returns are regressed on betas (Fama and French, 1992; Harvey and Siddique, 2000). The FF3F model typically results in R-squared values in excess of 0.70 (Ferguson and Shockley, 2002; Hahn and Lee, 2006; Petkova, 2006) while the quadratic CAPM produces R-squared values in excess of 0.6 (Harvey and Siddique, 2000). Importantly Harvey and Siddique (2000) and Dittmar (2002) respectively find that coskewness and cokurtosis remain significant and improve cross-sectional R-squared values when added to the FF3F model. This indicates that coskewness and cokurtosis explain cross-sectional variation in asset returns in excess of that explained by the FF3F model. Table 5.9 list cross-sectional adjusted R-squared values for the two sample periods across asset pricing models.

Table 5.9 Adjusted R-squared statistic for cross-sectional regressions

JSE listed share returns are sorted into portfolios for two sample periods: (1) January 1993 to December 2013; and (2) January 1997 to December 2013. Sample portfolios are formed using either independent decile sorts or two-way quintile sorts on size (ME) and BE/ME. The result is 10 portfolios formed on size, 10 portfolios formed on value and 25 portfolios formed on size and value. Finally industry portfolios are added to the sample for the January 1997 to December 2013 period. Mean portfolio returns are regressed on the respective factor coefficients for the following asset pricing models:

(A) CAPM	$E_T(R^{ei}) = \beta_i^m \lambda^m + \alpha^i$
(B) unconditional quadratic CAPM (CAPM^2)	$E_T(R^{ei}) = \beta_i^m \lambda^m + \beta_i^{m2} \lambda^{m2} + \alpha^i$
(C) unconditional cubic CAPM (CAPM^3)	$E_T(R^{ei}) = \beta_i^m \lambda^m + \beta_i^{m2} \lambda^{m2} + \beta_i^{m3} \lambda^{m3} + \alpha^i$
(D) FF3F model	$E_T(R^{ei}) = \beta_i^m \lambda^m + \beta_i^s \lambda^s + \beta_i^v \lambda^v + \alpha^i$
(E) APT	$E_T(R^{ei}) = \beta_i^m \lambda^m + \beta_i^s \lambda^s + \beta_i^v \lambda^v + \beta_i^{m2} \lambda^{m2} + \beta_i^{m3} \lambda^{m3} + \alpha^i$
(F) conditional quadratic CAPM (CCAPM^2)	$E_T(R^{ei}) = \beta_i^m \lambda^m + \beta_i^{sks} \lambda^{sks} + \alpha^i$
(G) conditional cubic CAPM (CCAPM^3)	$E_T(R^{ei}) = \beta_i^m \lambda^m + \beta_i^{sks} \lambda^{sks} + \beta_i^k \lambda^k + \alpha^i$
(H) CAPT	$E_T(R^{ei}) = \beta_i^m \lambda^m + \beta_i^s \lambda^v + \beta_i^v \lambda^v + \beta_i^{sks} \lambda^{sks} + \beta_i^k \lambda^k + \alpha^i$

The table lists the adjusted R-squared statistics of each cross-sectional regression.

January 1993 to December 2013

Panel	CAPM	CAPM^2	CAPM^3	FF3F	APT
Size	0.736	0.8	0.778	0.95	0.97
Value	0.8	0.854	0.942	0.968	0.986
Size-Value	0.678	0.668	0.711	0.939	0.947

January 1997 to December 2013

Panel	CAPM	CAPM^2	CAPM^3	FF3F	APT	CCCAPM^2	CCAPM^3	CAPT
Size	0.802	0.851	0.831	0.862	0.830	0.836	0.904	0.876
Value	0.862	0.892	0.976	0.962	0.993	0.877	0.921	0.97
Size-Value	0.706	0.694	0.735	0.828	0.838	0.699	0.686	0.992
Industry	0.752	0.732	0.745	0.756	0.629	0.716	0.813	0.7

The CAPM produces surprisingly high R-squared values for the sample period January 1993 to December 2013 ranging between 0.678 and 0.8 across panels. The unconditional quadratic CAPM increases the R-Squared from 0.736 to 0.8 for the size decile portfolios but is then reduced to 0.778 by the addition of the cubic market factor. This is in line with the results of the time-series regression indicating the cubic CAPM has little explanatory power for the expected returns of the size sorted portfolios. The FF3F model when estimated on the size panel produces an R-squared value of 0.95. This indicates that market, size and value factors are successful explaining almost all of the cross-sectional variation in expected return of the size sorted portfolios. The addition of coskewness and cokurtosis to the FF3F model produces only a marginally higher R-squared value which leaves one with the decision of whether the addition is warranted. The unconditional quadratic and cubic CAPM perform much better on the value decile portfolios increasing R-squared values from 0.8 to 0.854 and 0.942, respectively. This suggests that cubic CAPMs poor performance might be linked to the size sorted portfolio panel. The FF3F model produces a slightly higher R-squared value of 0.968 again proving its superiority. In contrast to the time-series results the cubic CAPM actually performs better than the quadratic CAPM on the size-value sorted portfolios but both fall substantially short of the FF3F model. The FF3F model would appear to be superior over the period January 1993 to December

2013 which is to be expected when recalling the concentration of the size and value premia between January 1993 and December 1996.

The results of the shorter sample period are similar to the January 1993 to December 2013 period for asset pricing models. Relative performance remains unchanged even if the exact R-squared values do not for the unconditional linear and higher moment CAPMs for the period January 1997 to December 2013. The quadratic CAPM increases the R-squared values on both size and value sorted panels while the cubic CAPM only performs well on the value sorted portfolios. The cubic CAPM produces a very high R-squared value of 0.976 for the value sorted portfolios. This suggests a link between book to market equity and cokurtosis. An APT model with FF3F and unconditional cubic CAPM factors produces an R-squared of 0.993 for the value sorted portfolios, effectively accounting for all cross-sectional variation in expected returns. The unconditional higher moment CAPMs again performs marginally better on the size-value sorted portfolio set. The FF3F model produces higher R-squared values for the size and size-value sorted portfolios but remarkably not on the value sorted portfolio panel.

The January 1997 to December 2013 sample period allows for the comparison of unconditional and conditional higher moment CAPM models. We see that only in one instance does a conditional model outperform an unconditional model when considering the size, value and size-value sorted panels. This leads us to conjecture that allowing for variation in coskewness and cokurtosis does not result in higher moment asset pricing models that better explain cross-sectional variation in asset returns. This is analysed in more depth in the next section.

Finally, industry portfolio mean returns are regressed on the factor betas of respective asset pricing models to provide an alternative to portfolios sorted on either market equity or BE / ME. Unconditional higher moments CAPMs underperform the linear CAPM while the FF3F model produces an R-squared value that is only 0.004 higher than the CAPM. This provides almost no justification for the addition of factors over the market factor when not regressing size or value sorted portfolios. The conditional cubic CAPM produces the highest R-squared value casting doubt on our earlier hypothesis regarding the superiority of unconditional models.

The results in Table 5.9 when viewed in their entirety do not lead to concrete conclusions regarding which asset pricing model best explains cross-sectional variation in expected returns. The perceived superiority of the FF3F model over the January 1993 to December 2013 sample period dissipates to some degree when the four years where the size and value premia are most concentrated is excluded. Furthermore, the relatively superior performance of the linear CAPM when regressing industry portfolio suggests that the ability of the higher moment CAPM and FF3F models might be limited to the ME and BE / ME sorted portfolio sets. The disparity in the results between portfolio sets highlights the dependence of results on the firm characteristic used in portfolio formation. Lastly no clear conclusion can be drawn regarding the superiority of unconditional or conditional model based on the results in this section.

Conditional vs Unconditional Asset Pricing Models

The central premise of conditional asset pricing models is that conditional alpha is always zero and that pricing errors result from the time variation in betas (Jensen, 1968; Jagannathan and Wang, 1996). It stands to reason then that the mean absolute pricing error of conditional models should be smaller than that of full information maximum likelihood models. Table 5.10 presents the mean absolute pricing error of FM regressions using constant and rolling betas for sample periods: (1) January 1993 to December 2013; and (2) January 1997 to December 2013. Size sorted, value sorted and size-value sorted portfolios are regressed on the betas of the respective asset pricing model. Industry portfolios are added to the sample for the shorter period. The rolling beta FM regression are completed using a window of 60 months as is customary. The implication is that investors are conditioned on five years of historic data. Conditional higher moment models are thus estimated in a different manner to that the rest of the study. The conditional model coefficients are estimated using a rolling regression technique rather than Harvey and Siddique's (2000) construction of a proxy portfolio.

Conditional mean absolute pricing errors are higher than their unconditional counterparts in all but three of the 15 values for the January 1993 to December 2013 sample period and three of the 20 values for the January 1997 to December 2013 sample period. These results do not support the hypothesis that conditional alpha is zero and furthermore indicate that unconditional models generally predict expected returns better. If an R-squared statistic characterises the ability of an asset pricing model to explain variation in expected returns then mean absolute pricing error characterises the accuracy with which an asset pricing model can predict historic returns. The FF3F model again appears superior to the linear and higher moment CAPMs for period January 1997 to December 2013 having the comparatively smallest mean absolute pricing error. The asset pricing model best able to predict historic return is an APT created by combining FF3F model factors with unconditional coskewness and cokurtosis. It should be kept in mind that the additional of unconditional coskewness and cokurtosis increases the R-squared of the FF3F model by between only one and two percent and as such is probably not justified.

The apparent superiority of the FF3F model does not hold as strongly over the sample period January 1997 to December 2013. The unconditional cubic CAPM outperforms the FF3F model on the value and industry panels while the linear and higher moment CAPMs outperform the FF3F model on the industry panel.

The results of this section corroborates the result of the portfolio sorts and time-series regressions. More importantly the conditional mean absolute pricing error is again higher than its unconditional counterpart in all but three instances. This result allows us to conclude with a reasonable degree of certainty that, on a portfolio by portfolios basis, conditional alpha is on average larger than unconditional alpha. This however is not a definitive conclusion that unconditional models outperform conditional models, one would need to test the hypothesis formally using a test statistic (e.g. Wald test) at a given confidence level to make such an assertion. This is not the focus of this study and as such we refer the discussion to the extensive literature on the subject e.g. Pettengill, Sundaram, and Mathur (1995), Lewellen and Nagel (2006) or Huang and Hueng (2008).

Table 5.10 Mean absolute pricing error for cross-sectional regressions

JSE listed share returns are sorted into portfolios for sample periods: (1) January 1993 to December 2013; and (2) January 1997 to December 2013. Sample portfolios are formed using either independent decile sorts or two-way quintile sorts on size (ME) and BE/ME. The result is 10 portfolios formed on size, 10 portfolios formed on value and 25 portfolios formed on size and value. Finally industry portfolios are added to the sample. Mean portfolio returns are regressed on the respective factor coefficients for the following asset pricing models:

(A) CAPM $E_T(R^{ei}) = \beta_i^m \lambda^m + \alpha^i$

(B) Quadratic CAPM (CAPM^2) $E_T(R^{ei}) = \beta_i^m \lambda^m + \beta_i^{m2} \lambda^{m2} + \alpha^i$

(C) Cubic CAPM (CAPM^3) $E_T(R^{ei}) = \beta_i^m \lambda^m + \beta_i^{m2} \lambda^{m2} + \beta_i^{m3} \lambda^{m3} + \alpha^i$

(D) FF3F model $E_T(R^{ei}) = \beta_i^m \lambda^m + \beta_i^s \lambda^s + \beta_i^v \lambda^v + \alpha^i$

(E) APT $E_T(R^{ei}) = \beta_i^m \lambda^m + \beta_i^s \lambda^s + \beta_i^v \lambda^v + \beta_i^{m2} \lambda^{m2} + \beta_i^{m3} \lambda^{m3} + \alpha^i$

The table lists the mean absolute pricing error of each regression. Portfolios are regressed using full information maximum likelihood FM regressions and rolling beta FM regressions.

	January 1993 to December 2013				
FM - constant betas	CAPM	CAPM^2	CAPM^3	FF3F	APT
Size	0.0034	0.0027	0.0028	0.0016	0.0011
Value	0.0031	0.0028	0.0018	0.0012	0.0006
Size-Value	0.0045	0.0045	0.0036	0.0019	0.0019
FM - rolling betas	CAPM	CAPM^2	CAPM^3	FF3F	APT
Size	0.0041	0.003	0.0023	0.0017	0.001
Value	0.0042	0.0033	0.0019	0.0011	0.0013
Size-Value	0.0054	0.005	0.0041	0.0023	0.002
	January 1997 to December 2013				
FMB - constant beta	CAPM	CAPM^2	CAPM^3	FF3F	APT
Size	0.0023	0.0019	0.0018	0.0016	0.0016
Value	0.002	0.0017	0.0009	0.001	0.0004
Size-Value	0.0031	0.0031	0.0026	0.0022	0.002
Industry	0.0033	0.003	0.0025	0.0029	0.0026
FMB - rolling beta	CAPM	CAPM^2	CAPM^3	FF3F	APT
Size	0.003	0.0028	0.0014	0.0024	0.0013
Value	0.0031	0.0036	0.0023	0.0012	0.0015
Size-Value	0.0039	0.0036	0.0034	0.0031	0.0026
Industry	0.0037	0.0019	0.0029	0.0043	0.002

Asset Pricing Model Risk Premia

Harvey and Siddique (2000) estimate risk premia for the FF3F model and FF3F model augmented by conditional coskewness using FM regressions for all shares (9,268) listed on the NYSE and AMEX between July 1993 and December 1993. They find risk premia for the market factor (λ^m), the value factor (λ^v) and the conditional coskewness factor (λ^{sks}) regression coefficients significant but interestingly the risk premium for the size factor is not significant. Cochrane (2005) suggests stripping all information in a regression not related to the factors when estimating factor risk premia. This idea makes intuitive sense and as such risk premia are estimated by cross-sectional regressions of expected excess returns with no intercept. Table 5.11 presents estimated factor risk premia of respective asset pricing models over the sample period January 1993 to December 2013. Risk premia are estimated for size sorted, value sorted and size-value sorted portfolio sets using full information maximum likelihood FM, OLS and GMM regressions. The standard errors of the coefficients are presented in parenthesis for each estimation technique.

The market risk premium (λ^m) is significant in all asset pricing models except the FF3F model and is generally quite large for the size sorted portfolios. This means that market beta is useful in explaining cross-sectional variation in portfolios formed on ME within the linear and higher moment CAPM models. The correlation between the market factor and SMB and HML are -0.35 and -0.22 respectively indicating that the market factor might not be significant due to collinearity. Alternatively the effects of the market factor might be subsumed by that of the SMB and HML. The market factor is further significant for both value sorted and size-value sorted portfolio in all asset pricing models except the APT. The market risk premium ranges between 0.6 and 1.3 percent per month across portfolios.

The risk premium associated with unconditional coskewness (λ^{m3}) is significant in size and value sorted portfolios but not in the size-value sorted portfolio set. Formally **H_1** is:

$$H_1: \lambda^{Skew} = 0$$

We therefore reject **H_1** at the 95 percent confidence level for the size sorted portfolio panel set and at the 90 percent confidence level for the value sorted portfolio set. We cannot reject **H_1** for the size-value sorted portfolio set. Unconditional cokurtosis is priced for size and value portfolio sets for the period January 1993 to December 2013 with the premium ranging between -0.5 and 0.41 percent per month. Unconditional coskewness thus helps explain the excess return associated with ME and BE / ME sorted portfolios. This confirms that a firm's market equity or size is related its coskewness beta.

Table 5.11 Factor risk premia per portfolio set (January 1993 to December 2013)

JSE listed share returns are sorted into portfolios for sample periods January 1993 to December 2013. Sample portfolios are formed using either independent decile sorts or two-way quintile sorts on size (ME) and BE/ME. The result is 10 portfolios formed on size, 10 portfolios formed on value and 25 portfolios formed on size and value. Finally industry portfolios are added to the sample. Mean portfolio returns are regressed on the respective factor coefficients for the following asset pricing models:

(A) CAPM $E_T(R^{ei}) = \beta_i^m \lambda^m + \alpha^i$

(B) unconditional quadratic CAPM (CAPM²) $E_T(R^{ei}) = \beta_i^m \lambda^m + \beta_i^{m2} \lambda^{m2} + \alpha^i$

(C) unconditional cubic CAPM (CAPM³) $E_T(R^{ei}) = \beta_i^m \lambda^m + \beta_i^{m2} \lambda^{m2} + \beta_i^{m3} \lambda^{m3} + \alpha^i$

(D) FF3F model $E_T(R^{ei}) = \beta_i^m \lambda^m + \beta_i^s \lambda^s + \beta_i^v \lambda^v + \alpha^i$

(E) APT $E_T(R^{ei}) = \beta_i^m \lambda^m + \beta_i^s \lambda^s + \beta_i^v \lambda^v + \beta_i^{m2} \lambda^{m2} + \beta_i^{m3} \lambda^{m3} + \alpha^i$

The table lists the regression coefficients combined with standard errors in parenthesis for different regression techniques FM, OLS and GMM. ***, **, * refer to the statistical significance of regression coefficients at the 1, 5 and 10% levels, respectively.

	CAPM	CAPM²		CAPM³		
	λ^m	λ^m	λ^{m2}	λ^m	λ^{m2}	λ^{m3}
Size	0.0128	0.0086	-0.0049	0.0088	-0.0050	0.0012
SE_{FM}	(0.0041)**	(0.0036)**	(0.0017)**	(0.0036)**	(0.0017)**	(0.0004)**
SE_{OLS}	(0.0042)**	(0.004)*	(0.002)**	(0.0041)*	(0.0026)*	(0.0005)*
SE_{GMM}	(0.0052)**	(0.004)*	(0.0026)	(0.0054)*	(0.0033)	(0.0007)**
Value	0.0095	0.0085	0.0041	0.0072	0.0041	-0.0013
SE_{FM}	(0.0037)**	(0.0036)**	(0.0019)*	(0.0036)*	(0.0019)*	(0.0005)**
SE_{OLS}	(0.0037)**	(0.0037)**	(0.0025)	(0.0037)*	(0.0025)	(0.0006)*
SE_{GMM}	(0.0039)**	(0.004)*	(0.0026)	(0.0052)*	(0.0033)	(0.0009)
Size - Value	0.0122	0.0117	-0.0010	0.0106	-0.0010	-0.0002
SE_{FM}	(0.0041)***	(0.0038)***	(0.0011)	(0.0038)**	(0.0011)	(0.0003)
SE_{OLS}	(0.0041)***	(0.0041)***	(0.0012)	(0.0042)**	(0.0013)	(0.0003)
SE_{GMM}	(0.0041)**	(0.0044)**	(0.0015)	(0.0042)**	(0.0016)	(0.0003)

	FF3F			5 Factor APT				
	λ^m	λ^s	λ^v	λ^m	λ^s	λ^v	λ^{m2}	λ^{m3}
Size	0.0062	0.0094	0.0330	0.0068	0.0016	0.0002	0.0086	0.0456
SE_{FM}	(0.0037)	(0.0041)*	(0.015)*	(0.0036)	(0.0018)	(0.0004)	(0.0039)*	(0.0163)**
SE_{OLS}	(0.0038)	(0.0042)*	(0.0166)*	(0.0037)	(0.0027)	(0.0005)	(0.004)*	(0.019)*
SE_{GMM}	(0.004)	(0.0049)*	(0.0215)	(0.0037)	(0.0037)	(0.0006)	(0.005)	(0.0352)
Value	0.0073	0.0125	0.0087	0.0069	0.0002	-0.0003	0.0102	0.0094
SE_{FM}	(0.0036)*	(0.0102)	(0.0043)*	(0.0036)	(0.0019)	(0.0005)	(0.0107)	(0.0043)*
SE_{OLS}	(0.0036)*	(0.0105)	(0.0043)*	(0.0037)	(0.0023)	(0.0005)	(0.0117)	(0.0044)*
SE_{GMM}	(0.0036)*	(0.0125)	(0.0053)	(0.0036)	(0.002)	(0.0006)	(0.0141)	(0.0055)
Size - Value	0.0066	0.0082	0.0120	0.0070	0.0012	0.0000	0.0081	0.0104
SE_{FM}	(0.0037)*	(0.0041)*	(0.0052)**	(0.0037)*	(0.0012)	(0.0003)	(0.0041)*	(0.0048)**
SE_{OLS}	(0.0038)*	(0.0041)*	(0.0052)**	(0.0037)*	(0.0015)	(0.0004)	(0.0041)*	(0.005)*
SE_{GMM}	(0.0186)*	(0.038)**	(0.0445)**	(0.0039)*	(0.0015)	(0.0004)	(0.0046)*	(0.0058)*

The risk premium associated with unconditional cokurtosis (λ^{m3}) is significant in size and value sorted portfolios but not in the size-value sorted portfolio set. Formally **H_2** is:

$$H_2: \lambda^{Kurt} = 0$$

We therefore reject **H_2** at the 95 percent confidence level for the size sorted portfolio panel set and at the 95 percent confidence level for the value sorted portfolio set. We cannot reject **H_2** for the size-value sorted portfolio set. Unconditional cokurtosis is thus priced for size and value portfolio sets for the period January 1993 to December 2013 but the premium is never very large ranging between -0.13 and 0.12 percent per month.

The risk premia for the SMB and HML factors are significant for size sorted and size-value sorted portfolios but only HML is significant for the value sorted portfolio panel. The risk premia are generally large ranging between 0.8 and 1.3 percent with one exception viz. λ^{v} is 3.3 percent per month for the size sorted portfolio panel. The SMB and HML factors are thus important in explaining the variation in expected returns of size sorted and size-value sorted portfolios.

The combination of the FF3F and unconditional cubic CAPM models into an APT model produces unexpected results. Only coskewness and cokurtosis are priced for the size portfolio set. The risk premia on coskewness and cokurtosis actually increase in magnitude becoming 0.9 and 4.6 percent per month respectively. The excess return associated with SMB and HML factors are thus absorbed by coskewness and cokurtosis. Only cokurtosis is significant for the value sorted portfolios panel.

The standard errors of the FM regression are in most cases marginally smaller than that of the OLS regressions and substantially smaller than the GMM standard errors. As all these techniques produce the same regression coefficients, t-statistics will be highest for the FM regressions in most cases. The FM regression as constructed in this study accommodates for autocorrelation only in the cross-section and not in the time-series whereas the GMM does. The GMM coefficient might thus be insignificant because of autocorrelation in the time-series.

Table 5.12 presents estimated factor risk premia of respective asset pricing models for the sample period January 1997 to December 2013. Risk premia are estimated for size sorted, value sorted, size-value sorted and industry sorted portfolios using full information maximum likelihood FM, OLS and GMM regressions. The standard errors of the coefficients are presented in parenthesis for each estimation technique.

Table 5.12 Factor risk premia per portfolio set (January 1997 to December 2013)

JSE listed share returns are sorted into portfolios for sample periods January 1997 to December 2013. Sample portfolios are formed using either independent decile sorts or two-way quintile sorts on size (ME) and BE/ME. The result is 10 portfolios formed on size, 10 portfolios formed on value and 25 portfolios formed on size and value. Mean portfolio returns are regressed on the respective factor coefficients for the following asset pricing models:

(A) CAPM $E_T(R^{ei}) = \beta_i^m \lambda^m + \alpha^i$
(B) unconditional quadratic CAPM (CAPM^2) $E_T(R^{ei}) = \beta_i^m \lambda^m + \beta_i^{m2} \lambda^{m2} + \alpha^i$
(C) unconditional cubic CAPM (CAPM^3) $E_T(R^{ei}) = \beta_i^m \lambda^m + \beta_i^{m2} \lambda^{m2} + \beta_i^{m3} \lambda^{m3} + \alpha^i$
(D) conditional quadratic CAPM (CCAPM^2) $E_T(R^{ei}) = \beta_i^m \lambda^m + \beta_i^{sks} \lambda^{sks} + \alpha^i$
(E) conditional cubic CAPM (CCAPM^3) $E_T(R^{ei}) = \beta_i^m \lambda^m + \beta_i^{sks} \lambda^{sks} + \beta_i^k \lambda^k + \alpha^i$
(F) FF3F model $E_T(R^{ei}) = \beta_i^m \lambda^m + \beta_i^s \lambda^s + \beta_i^v \lambda^v + \alpha^i$

The table lists the regression coefficients combined with standard errors in parenthesis for different regression techniques FM, OLS and GMM. ***, **, * refer to the statistical significance of regression coefficients at the 1, 5 and 10% levels, respectively.

	CAPM	CAPM^2		CAPM^3		
Panel	λ^m	λ^m	λ^{m2}	λ^m	λ^{m2}	λ^{m3}
Size	0.0094	0.0069	-0.0025	0.0070	-0.0025	0.0007
SE_{FM}	(0.0047)*	(0.0042)	(0.0016)	(0.0042)	(0.0016)	(0.0004)
SE_{OLS}	(0.0047)*	(0.0042)	(0.0021)	(0.0043)	(0.0023)	(0.0005)
SE_{GMM}	(0.0059)	(0.0044)	(0.0021)	(0.0041)	(0.0021)	(0.0004)
Value	0.0084	0.0076	0.0024	0.0070	0.0030	-0.0010
SE_{FM}	(0.0042)*	(0.0042)	(0.0018)	(0.0042)	(0.0019)	(0.0006)
SE_{OLS}	(0.0042)*	(0.0042)	(0.0023)	(0.0042)	(0.0025)	(0.0007)
SE_{GMM}	(0.0043)*	(0.0043)	(0.0023)	(0.0042)	(0.0027)	(0.0009)
Size - Value	0.0087	0.0088	-0.0002	0.0083	-0.0001	-0.0002
SE_{FM}	(0.0046)*	(0.0043)*	(0.0012)	(0.0043)*	(0.0012)	(0.0003)
SE_{OLS}	(0.0046)*	(0.0044)*	(0.0013)	(0.0044)*	(0.0013)	(0.0003)
SE_{GMM}	(0.0055)	(0.0046)*	(0.0017)	(0.0044)*	(0.0016)	(0.0003)
Industry	0.0085	0.0080	-0.0018	0.0074	-0.0052	0.0003
SE_{FM}	(0.0045)	(0.0044)	(0.0022)	(0.0045)	(0.0034)	(0.0005)

SE_{OLS}	(0.0045)		(0.0045)	(0.0028)		(0.0055)	(0.0088)	(0.001)	
SE_{GMM}	(0.0045)		(0.0043)	(0.0028)		(0.0067)	(0.0096)	(0.0017)	
	CCAPM[2]			CCAPM[3]				FF3F	
Panel	λ^m	λ^{sks}		λ^m	λ^{sks}	λ^k		λ^m	λ^s	λ^v
Size	0.0062	0.0146		0.0055	0.0202	0.0223		0.0062	0.0021	0.0137
SE_{FM}	(0.0041)	(0.007)*		(0.0041)	(0.008)**	(0.0094)**		(0.0042)	(0.0038)	(0.017)
SE_{OLS}	(0.0041)	(0.0073)*		(0.0042)	(0.0111)	(0.0125)		(0.0042)	(0.0038)	(0.0169)
SE_{GMM}	(0.0044)	(0.0093)		(0.0044)	(0.013)	(0.0154)		(0.0042)	(0.0044)	(0.0223)
Value	0.0100	-0.0010		0.0121	-0.0060	0.0024		0.0073	0.0048	0.0049
SE_{FM}	(0.0044)*	(0.0073)		(0.0047)**	(0.0085)	(0.0069)		(0.0042)	(0.0117)	(0.0048)
SE_{OLS}	(0.0044)*	(0.0078)		(0.0049)**	(0.0098)	(0.0078)		(0.0042)	(0.012)	(0.0048)
SE_{GMM}	(0.0045)*	(0.0076)		(0.0053)*	(0.0099)	(0.0085)		(0.0041)	(0.0133)	(0.0056)
Size - Value	0.0076	0.0097		0.0075	0.0102	0.0091		0.0067	0.0001	0.0105
SE_{FM}	(0.0042)*	(0.0066)		(0.0042)*	(0.0069)	(0.0058)		(0.0067)	(0.0001)	(0.0105)*
SE_{OLS}	(0.0043)*	(0.0067)		(0.0043)*	(0.007)	(0.0058)		(0.0066)	(0.0001)	(0.0067)*
SE_{GMM}	(0.0046)	(0.009)		(0.0047)	(0.0093)	(0.0065)		(0.0053)	(0.0012)	(0.0104)*
Industry	0.0079	0.0082		0.0081	0.0063	0.0002		0.0063	0.0095	0.0092
SE_{FM}	(0.0043)	(0.0052)		(0.0043)	(0.0054)	(0.0067)		(0.0045)	(0.0106)	(0.0116)
SE_{OLS}	(0.0043)	(0.0053)		(0.0044)	(0.0056)	(0.0069)		(0.0046)	(0.0111)	(0.0121)
SE_{GMM}	(0.0044)	(0.0057)		(0.0044)	(0.0058)	(0.0072)		(0.0044)	(0.0117)	(0.0157)

The results of the January 1997 to December 2013 sample period are very disappointing. The coefficient of the unconditional higher moment CAPMs and the FF3F model are for the largest part no longer significant. This means that unconditional coskewness, unconditional cokurtosis, SMB and HML are no longer priced and are not important in explaining cross-sectional variation in expected returns of the various portfolios. We can therefore not reject $\mathbf{H_1}$ and $\mathbf{H_2}$ for unconditional higher moment models. There are three possible explanations. This can either be as a result of the residual covariance matrix used in the regressions or because the decreased length of the time-series, both of which can increase the standard error of the estimates. Neither reason seems likely as standard errors are reasonably stable compared to the longer sample period. The more likely cause of the t-statistics deteriorating is the substantial reduction in the magnitude of factor risk premia. Recalling that the magnitude of the excess return on the size and value sorted portfolios are no longer significant over the January 1997 to December 2013 sample period, it should not be surprising that factors no longer have risk premia associated with them.

Interestingly, conditional coskewness and cokurtosis are statistically significant for the size sorted portfolio panel and are generally quite large ranging between 1.6 and 2.3 percent per month. We can therefore reject $\mathbf{H_1}$ at the 90 percent confidence level and $\mathbf{H_2}$ at the 95 percent confidence level for conditional higher moment models. Conditional coskewness and cokurtosis with the market portfolio are thus priced when portfolios are formed on ME over the sample period 1997 to December 2013.

Not a single asset pricing model coefficient is significant for the industry portfolio set again cast doubt on the ability of the chosen model factors to explain variation in expected return in portfolios not formed on market equity or BE / ME.

Expected Return, Market Beta, Coskewness Beta and Cokurtosis Beta

The intuitive appeal of a positive linear relationship between expected return and systematic risk cannot be denied. A fundamental axiom of the CAPM is that an investor can only earn a higher return by accepting more risk i.e. invest in high beta stocks. The study covered a large body of literature suggesting that this is not the case. Contrary to economic and utility theory Van Rensburg and Robertson (2003) find an inverse relationship between market beta and expected return on the JSE. This section seeks to confirm this finding as well as extending the analysis to coskewness beta and cokurtosis beta.

In order to isolate the relationship between expected return with market beta, coskewness beta or cokurtosis beta, sample portfolio means are individually regressed on the time-series regression coefficients of the market factor, the market factor squared and the market factor cubed. The FM explanatory variables are thus estimated form single factors regressions rather than the multiple regressions used earlier in the study. It is important to note that cross-sectional regressions are specified with an intercept and that the dependent variable is the expected return not expected excess return of sample portfolios. The cross-sectional regression intercepts should thus be equal to the risk free rate or return on the zero beta portfolio. Regressions are run in this manner because in the alternative where expected excess returns are regressed on coefficients, one effectively forces the cross-sectional regression intercept to zero. Cochrane (2005) indicates that removing the risk-free rate

from dependent variables is desirable when estimating risk premia or testing if pricing errors are zero but forcing the intercept to zero could distort the nature of the relationship between dependent and independent variables. The following cross-sectional regressions are specified:

$$E_T(R^i) = \lambda^0 + \beta_i^m \lambda^m + \alpha^i \qquad i = 1, \dots, N \qquad (5.16)$$

$$E_T(R^i) = \lambda^0 + \beta_i^{m2} \lambda^{m2} + \alpha^i \qquad i = 1, \dots, N \qquad (5.17)$$

$$E_T(R^i) = \lambda^0 + \beta_i^{m3} \lambda^{m3} + \alpha^i \qquad i = 1, \dots, N \qquad (5.18)$$

The regressions are illustrated in scatter plots rather than represented in table format to facilitate graphical interpretation (Refer to Figure 5.1).

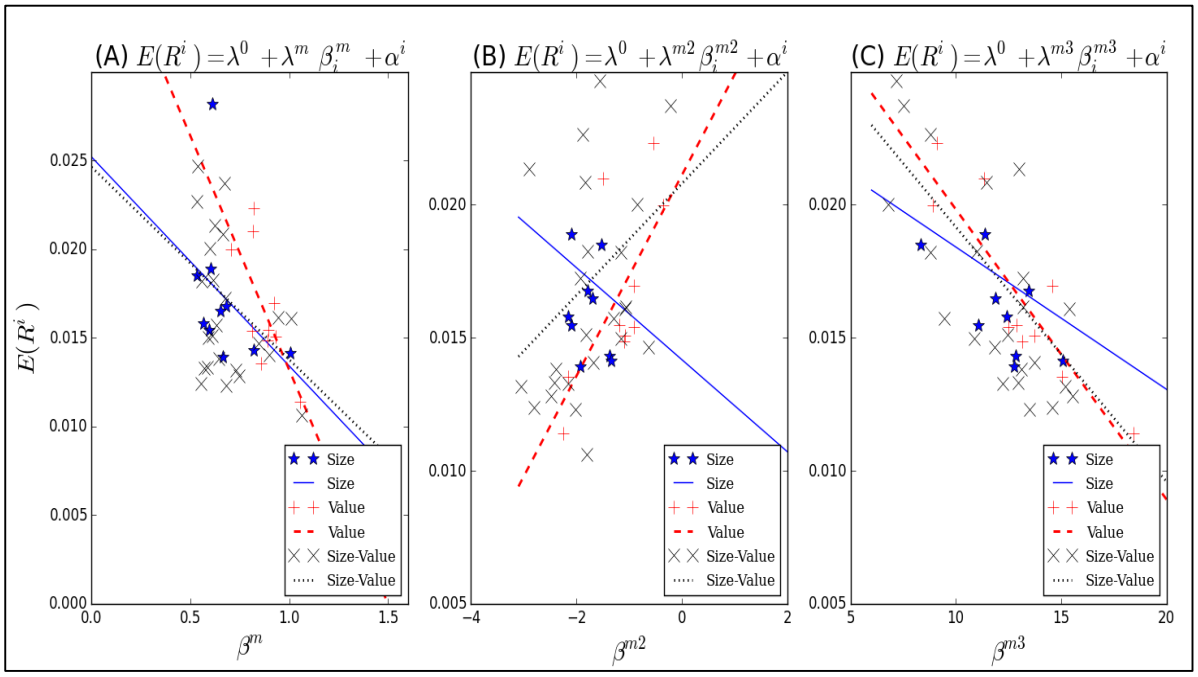


Figure 5.1 Scatter plots of expected return vs market beta, coskewness beta and cokurtosis beta

Figure 5.1 (A) shows a clear inverse or negative relationship between expected return and market beta. The study therefore confirms Van Rensburg and Robertson's (2003) result of an inverse relationship between expected return and beta. It must be noted that this result completely invalidates the central premise of the CAPM, a positive linear relationship between risk and return.

The results within Figure 5.1 (B) are not as clear. Expected return is negatively related to coskewness beta for the size sorted portfolio set but the relationship is positive for the value sorted and size-value sorted portfolio sets. Harvey and Siddique's (2000) hypothesise a negative relationship based on investor non-increasing absolute risk aversion. The dispersion of the portfolios formed on size suggest that the negative relationship between size and expected return might be as a result of a poor model fit.

Dittmar (2002) hypothesises a positive relationship between expected return and cokurtosis beta based on a prudent investor's aversion to extreme outcomes. The results within Figure 5.1 (C) clearly show a negative relationship between expected return and cokurtosis beta. This means that investors

would be rewarded for holding stocks that are less susceptible to extreme market movements. This counterintuitive argument is similar to Van Rensburg and Robertson's (2003) finding that investors obtain higher returns by investing in lower risk stocks. However, the poor performance of cokurtosis beta within the time-series regressions suggests that this result is far from a definitive conclusion.

Asset Pricing Model Tests

Harvey and Siddique (2000) complete F-tests to determine if intercepts are jointly zero on eight different portfolio panels including the size sorted, value sorted and size-value sorted sets. They find that the addition of conditional coskewness to the FF3F model reduces the F-statistic dramatically but still strongly rejects **H_3** for size sorted, value sorted and size-value sorted portfolio panels. Vorkink (2003) employs time-series test statistic based on both OLS and GMM estimation on size and momentum sorted portfolios. He rejects the linear CAPM, conditional higher moment CAPMS and FF3F model. Basiewicz and Auret (2010) utilise time-series GRS test statistics on size-value sorted portfolios to establish the mean-variance efficiency of the CAPM, the APT proposed by Van Rensburg and Slaney (1997) and FF3F models. On the value-weighted portfolios they reject the CAPM at the ten percent significance level, the APT at the five percent significance level but cannot reject the FF3F model.

As discussed previously we utilise a cross-sectional or two-pass testing methodology employing OLS, GLS, FM and GMM test statistics to assess the linear CAPM, unconditional and conditional higher moment CAPMs and finally the FF3F model. Formally **H_3** is defined as:

$$H_3: \hat{\alpha} = 0$$

Table 5.13 lists test statistics J_{OLS}, J_{GLS}, J_{FM} and J_{GMM} of respective asset pricing models for sample periods: (1) January 1993 to December 2013; and (2) January 1997 to December 2013. Test are estimated for size sorted, value sorted, size-value sorted and industry sorted portfolios using full information maximum likelihood regressions. The corresponding p-values of the test statistics are presented in parenthesis.

For the sample period January 1993 to December 2013 the linear CAPM, the unconditional quadratic CAPM and the unconditional cubic CAPM produce significant test statistics (J_{OLS}, J_{GLS}, J_{FM} and J_{GMM}) for the size sorted portfolio set while the FF3F model does not. We therefore reject **H_3** for the CAPM, the unconditional quadratic CAPM and the unconditional cubic CAPM for the size portfolio set. This indicates that the linear and higher moment CAPM models do not accurately quantify the relationship between risk and expected return. As the FF3F model is not rejected one can conclude that it does price size sorted portfolios accurately over the sample period. The study therefore cannot reject the hypothesis that the FF3F model accurately prices assets for the size sorted portfolio set.

Interestingly, none of the test statistics are significant for the value sorted portfolio panel so we cannot reject **H_3**. This study can thus not reject the hypothesis that the asset pricing models tested accurately price assets. The cubic CAPM and FF3F models produce lower test statistics than the CAPM and quadratic CAPM indicating that pricing errors are lower for the value sorted portfolio set. The size-value sorted portfolios test statistics are not as uniform in their conclusion as in previous two panels.

J_{OLS} is not significant for the CAPM while J_{GLS}, J_{FM} and J_{GMM} are significant. As the majority of the test statistics are significant, we can only reject $\mathbf{H_3}$ for the CAPM. Only J_{GMM} is significant for the unconditional higher moment CAPMs and the FF3F model. Consistency therefore demands that we do not reject $\mathbf{H_3}$.

The January 1997 to December 2013 sample period facilitates the inclusion of conditional higher moment models as well as the industry portfolio panel. Even though test statistic values of portfolio panels differ from the test statics of longer sample period, the conclusions drawn from the test statistics do not except in a single case viz. the FF3F model on size sorted portfolios. As such the discussion will focus on conditional models, the industry panel and the FF3F model.

Test statistics J_{OLS}, J_{GLS}, J_{FM} and J_{GMM} for the conditional quadratic CAPM are significant when evaluating the size portfolio set. We thus reject $\mathbf{H_3}$ when estimating conditional quadratic CAPM using the size portfolio set. Only J_{GLS} is significant for conditional cubic CAPM and as such $\mathbf{H_3}$ should not be rejected. J_{GLS} and J_{GMM} are significant on the size portfolio for the FF3F model which again casts doubts about its superiority over the shorter sample period.

Test statistics in the main are not significant for the industry portfolio panel except the FF3F model. We therefore cannot reject $\mathbf{H_3}$ for the linear CAPM, unconditional quadratic CAPM, unconditional cubic CAPM, conditional quadratic CAPM and the conditional cubic CAPM. The test statistics J_{OLS}, J_{GLS}, J_{FM} and J_{GMM} on the industry panel for the FF3F model are significant. We thus for the first time reject $\mathbf{H_3}$ for the FF3F model on the industry portfolio set.

The disparity in the results of the test statistics again highlights the dependence of results on the firm characteristic used in portfolio formation. The selection of only the size sorted or value sorted portfolio panels as the sample data would have resulted in completely different results. $\mathbf{H_3}$ would have been rejected for all asset pricing models barring the FF3F model should sample data purely consist of size sorted portfolios and conversely $\mathbf{H_3}$ would not have been rejected for a single asset pricing model if value sorted portfolio if sample data purely consisted of value sorted portfolio.

Table 5.13 Asset pricing model tests

JSE listed share returns are sorted into portfolios for two sample periods: (1) January 1993 to December 2013; and (2) January 1997 to December 2013. Sample portfolios are formed using either independent decile sorts or two-way quintile sorts on size (ME) and BE/ME. The result is 10 portfolios formed on size, 10 portfolios formed on value and 25 portfolios formed on size and value. Finally industry portfolios are added to the sample for the January 1997 to December 2013 period. Mean portfolio returns are regressed on the respective factor portfolios for the following asset pricing models:

(A) CAPM $E_T(R^{ei}) = \beta_i^m \lambda^m + \alpha^i$
(B) unconditional quadratic CAPM (CAPM^2) $E_T(R^{ei}) = \beta_i^m \lambda^m + \beta_i^{m2} \lambda^{m2} + \alpha^i$
(C) unconditional cubic CAPM (CAPM^3) $E_T(R^{ei}) = \beta_i^m \lambda^m + \beta_i^{m2} \lambda^{m2} + \beta_i^{m3} \lambda^{m3} + \alpha^i$
(D) FF3F model $E_T(R^{ei}) = \beta_i^m \lambda^m + \beta_i^s \lambda^s + \beta_i^v \lambda^v + \alpha^i$
(E) conditional quadratic CAPM (CCAPM^2) $E_T(R^{ei}) = \beta_i^m \lambda^m + \beta_i^{sks} \lambda^{sks} + \alpha^i$
(F) conditional cubic CAPM (CCAPM^3) $E_T(R^{ei}) = \beta_i^m \lambda^m + \beta_i^{sks} \lambda^{sks} + \beta_i^k \lambda^k + \alpha^i$

The table lists test statistics J_{OLS}, J_{GLS}, J_{FMB}, and J_{GMM} and respective probabilities. ***, **, * refer to the rejection of the hypothesis that pricing errors are jointly zero at the 1, 5 and 10% significance levels, respectively.

	(1) January 1993 to December 2013							
	CAPM		CAPM^2		CAPM^3		FF3F	
Size								
J_{OLS}	20.2**	(0.017)	18**	(0.021)	18**	(0.012)	8.4	(0.299)
J_{GLS}	20.8**	(0.014)	21.1***	(0.007)	20.5***	(0.005)	9.5	(0.221)
J_{FM}	21.1**	(0.012)	18.8**	(0.016)	18.8***	(0.009)	8.8	(0.269)
J_{GMM}	22***	(0.009)	16.2**	(0.04)	17.1**	(0.017)	10.5	(0.161)
Value								
J_{OLS}	11.3	(0.258)	10.5	(0.232)	4.4	(0.727)	2.9	(0.891)
J_{GLS}	11.6	(0.237)	9.7	(0.287)	6.6	(0.471)	2.9	(0.892)
J_{FM}	11.8	(0.226)	11	(0.204)	4.6	(0.703)	3.1	(0.879)
J_{GMM}	5.7	(0.766)	5.3	(0.724)	3.2	(0.861)	2	(0.96)
Size-value								
J_{OLS}	31.7	(0.134)	27.8	(0.222)	25.7	(0.266)	21.2	(0.511)
J_{GLS}	34.8*	(0.071)	26.8	(0.266)	27.3	(0.2)	22.4	(0.438)
J_{FM}	35.4*	(0.063)	31	(0.122)	28.6	(0.156)	23.6	(0.369)
J_{GMM}	52.3***	(0.001)	44.1***	(0.005)	39.8**	(0.011)	59.1***	(0)

	(2) January 1997 to December 2013											
	CAPM		$CAPM^2$		$CAPM^3$		$CCAPM^2$		$CCAPM^3$		FF3F	
Size												
J_{OLS}	17.6**	(0.04)	17**	(0.03)	17**	(0.017)	14.9*	(0.062)	10.7	(0.151)	10	(0.187)
J_{GLS}	18.4**	(0.031)	18.4**	(0.018)	19.1***	(0.008)	17.3**	(0.027)	17.5**	(0.014)	12.6*	(0.081)
J_{FM}	18.6**	(0.028)	18**	(0.021)	18**	(0.012)	15.7**	(0.047)	11.3	(0.125)	10.6	(0.157)
J_{GMM}	22.8***	(0.007)	21.7***	(0.006)	23.1***	(0.002)	19.3**	(0.013)	10.4	(0.165)	13.1*	(0.069)
Value												
J_{OLS}	4.7	(0.859)	3.8	(0.873)	1.1	(0.992)	4.2	(0.836)	3.6	(0.825)	1.9	(0.967)
J_{GLS}	4.9	(0.842)	3.5	(0.901)	1.8	(0.971)	4.3	(0.828)	3.8	(0.801)	1.9	(0.964)
J_{FM}	5	(0.836)	4	(0.854)	1.2	(0.991)	4.5	(0.812)	3.8	(0.803)	2	(0.962)
J_{GMM}	2.8	(0.971)	2.5	(0.96)	1	(0.995)	3	(0.937)	2.7	(0.913)	1.2	(0.991)
Size-value												
J_{OLS}	30.1	(0.182)	26.6	(0.272)	22.7	(0.419)	29.7	(0.157)	29.5	(0.131)	26.5	(0.232)
J_{GLS}	34*	(0.085)	26.2	(0.291)	27.3	(0.202)	34.1*	(0.063)	34.2**	(0.047)	29.6	(0.129)
J_{FM}	34.5*	(0.077)	30.5	(0.135)	26	(0.251)	34.1*	(0.064)	33.8*	(0.051)	30.3	(0.111)
J_{GMM}	57.3***	(0)	52.8***	(0)	51.9***	(0)	56.9***	(0)	57.6***	(0)	63.6***	(0)
Industry												
J_{OLS}	10.1	(0.184)	10.1	(0.122)	7.4	(0.194)	10.1	(0.122)	7.9	(0.164)	9.3*	(0.097)
J_{GLS}	10.4	(0.167)	10.2	(0.117)	18.5***	(0.002)	10.4	(0.11)	8.2	(0.144)	9.6*	(0.088)
J_{FM}	10.5	(0.16)	10.5	(0.104)	7.7	(0.172)	10.5	(0.104)	8.2	(0.145)	9.7*	(0.083)
J_{GMM}	12.4*	(0.088)	11*	(0.089)	3.9	(0.563)	11.1*	(0.086)	8.1	(0.15)	11.2**	(0.047)

5.5 Monte Carlo Results

The test statistics in the previous section lacked consistency over estimation techniques. J_{OLS}, J_{GLS}, J_{FM} and J_{GMM} at times resulted in differing conclusions. DM (1998) provide graphical tools that assess the size and power characteristics of test statistics. The DM (1998) tools are based on a Monte Carlo simulation methodology where outcomes are known and test statistic values are bootstrapped. The DM (1998) p-value plots and size-power curves should provide the necessary guidance we required in order to give due consideration to the test statistics utilised in the previous section. It is important to note that for the purposes of this section T refers to the length of the time series while N refers to the number portfolios and K to the number of factors used in each simulation.

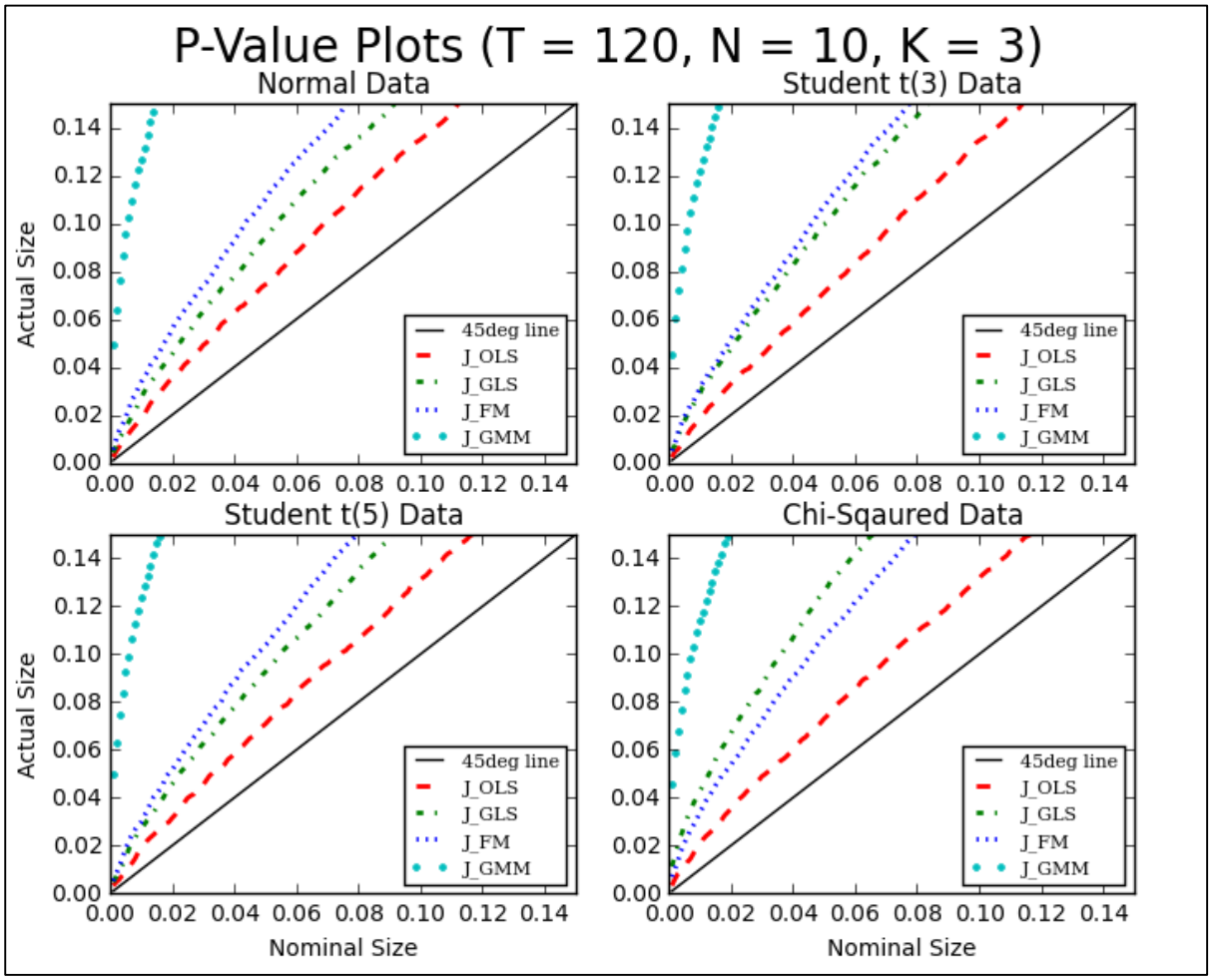


Figure 5.2 P-value plots where T = 120

The simulation is completed as described in Section 4.5 and inputs are varied to determine their impact on test statistic values. The length of the time-series is varied to 120, 180 and 252 in order to determine the impact of sample size. The number of factors do not affect specific test statistics on an isolated basis and as such is kept constant i.e. changes the p-value or size-power curves uniformly across test statistics. Figures display p-value plots or size-power curves for J_{OLS}, J_{GLS}, J_{FM} and J_{GMM} for data simulated with errors drawn from: (a) a normal distribution; (b) a central $t_{(3)}$ distribution; (c) a central $t_{(5)}$ distribution; and (d) a $\chi^2_{(4)}$ distributions. Errors are drawn from the central $t_{(3)}$ and central

$t_{(5)}$ distributions to investigate the effect of varying degrees of kurtosis while errors are drawn from the $\chi^2_{(4)}$ distribution to investigate the effect of skewness on the test statistics. P-value plots are scaled to 0.15 as hypotheses are rejected at the ten percent level.

Figure 5.1 displays the p-value plots for J_{OLS}, J_{GLS}, J_{FM} and J_{GMM} when $T = 120$. Recall that if the p-value plot is above (below) the 45° line, the test-statistic systematically over (under) rejects the null hypothesis. All test statistics over reject the null hypothesis but J_{OLS} performs best and J_{GMM} worst regardless of the distributional nature of the data. The performance of J_{GMM} is exceptionally poor. Figure 5.1 indicates that J_{GMM} incorrectly rejects the null the majority of the time. Furthermore J_{GLS} performs better than J_{FM} in all cases except when data is skewed. The relative poor performance of the of all test-statistics in Figure 5.1 clearly illustrates why in finance more than 120 data points should be used, particularly when using a GMM methodology.

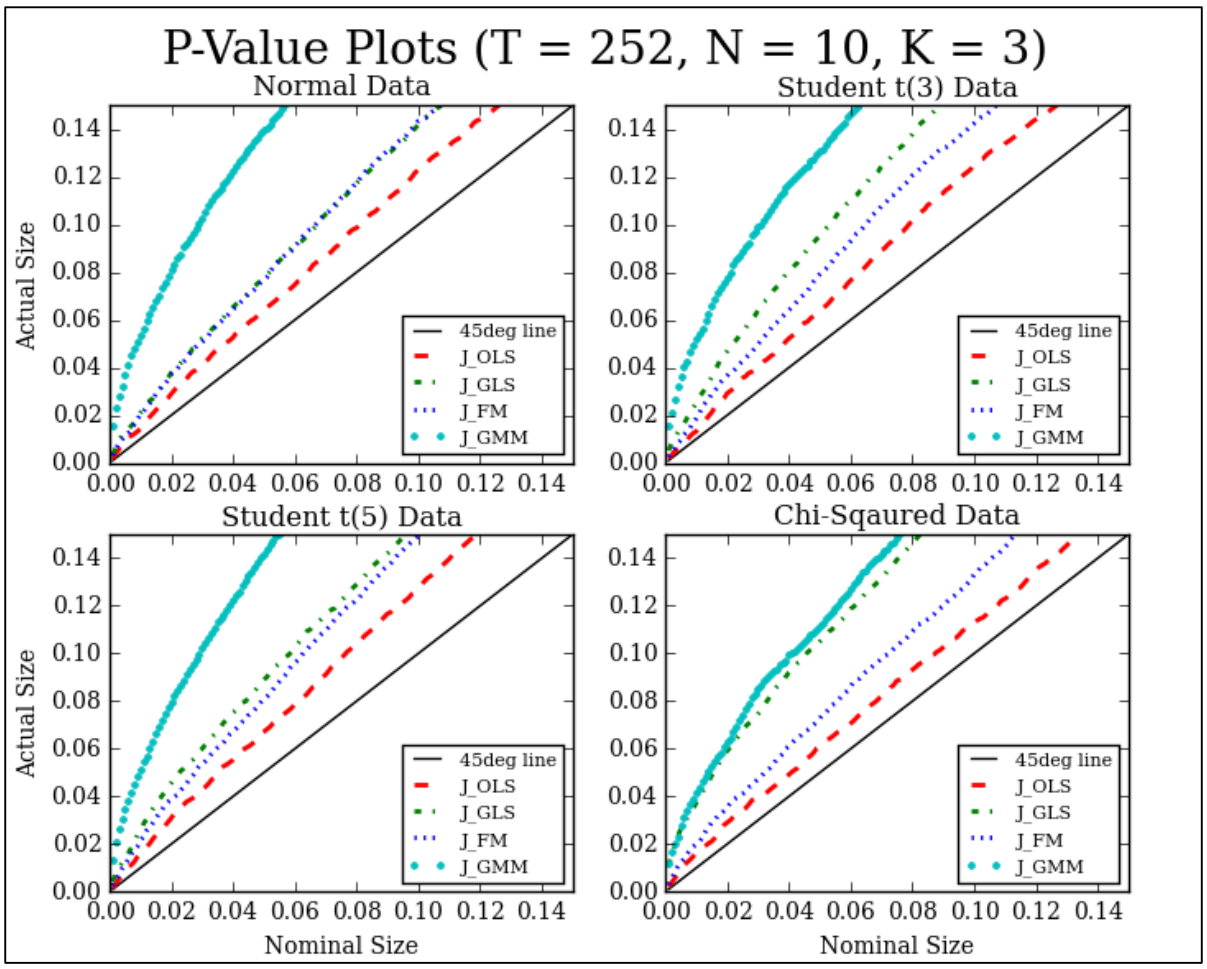


Figure 5.3 P-value plots where T = 252

The p-value plot where $T = 180$ conveys much the same information as the p-value plot for $T = 252$ and as such only the latter is displayed. As can be expected due the longer time-series Figure 5.2 shows slight improvements for J_{OLS}, J_{GLS}, and J_{FM}. The disproportionate improvement in the performance of the J_{GMM} in Figure 5.2 is unexpected. The J_{GMM} is still the worst performing test statistic by a reasonable margin and as the length of the time-series in the paper is 252 data points we should be very careful of accepting the conclusion from J_{GMM} when different from other test statistics. J_{OLS} again performs

best regardless of the distributional nature of the data and thus should be used when presented with differing conclusions. Interestingly J_{OLS} and J_{FM} seem to perform better when skewness is introduced into the data. Lastly it should be noted that J_{GLS} no longer performs better than J_{FM} which almost defines a pecking order for the statistic to be used based on size characteristics.

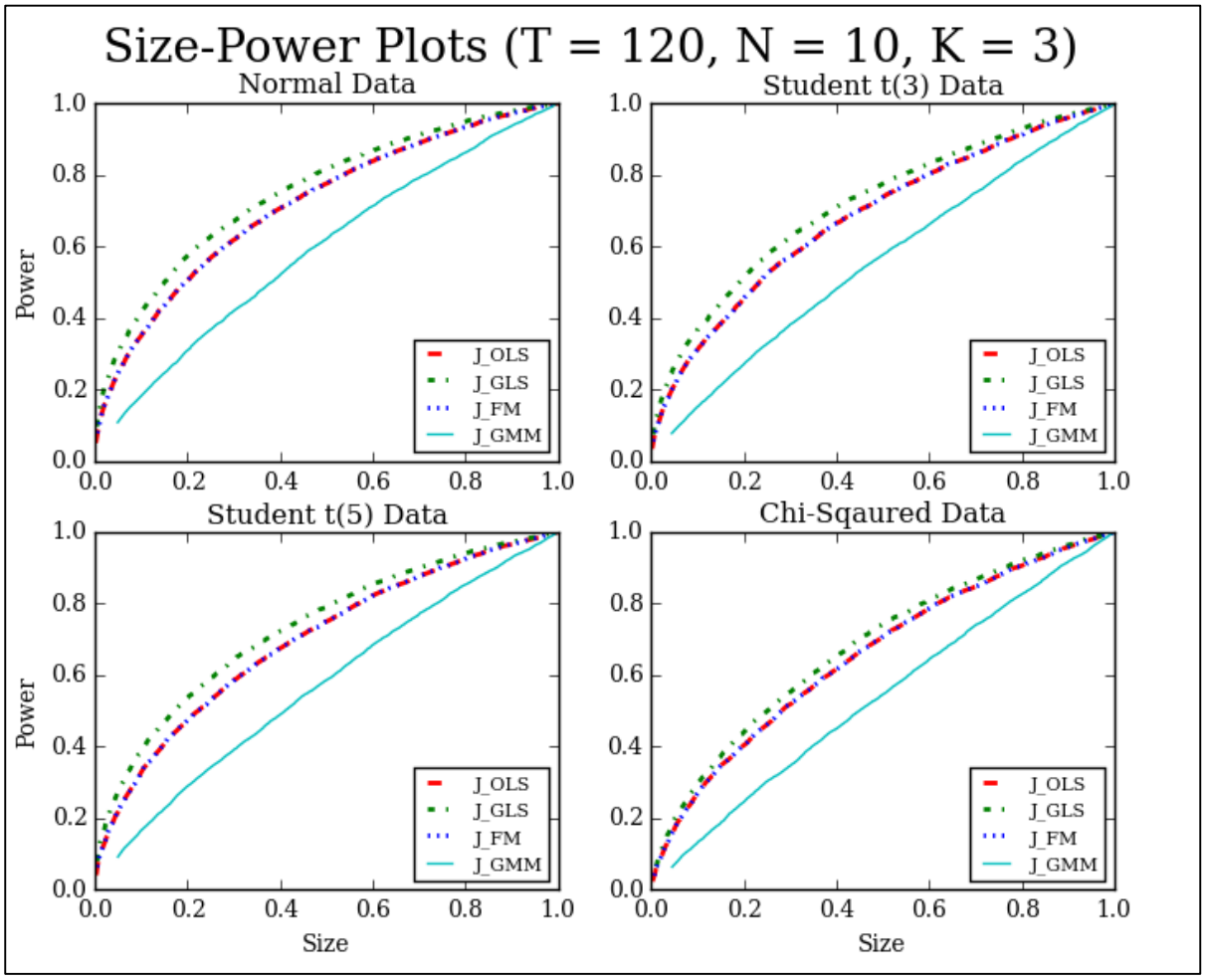


Figure 5.4 Size-power plots where T = 120

Figure 5.3 displays the size-power curves for J_{OLS}, J_{GLS}, J_{FM} and J_{GMM} when $T = 120$. Recalling that test statistics have greater power, that is they correctly reject the alternative hypothesis more often, if their curves are above the other statistics simulated. J_{GMM} again performs poorly indicating that the alternative hypothesis is not rejected as it should be. J_{OLS} and J_{FM} have identical power properties but as Cochrane (2005) points out J_{GLS} is the most powerful of the test statistics.

Figure 5.4 displays the size-power curves for J_{OLS}, J_{GLS}, J_{FM} and J_{GMM} when $T = 252$. As is to be expected all test statistics have greater power when the length of the time-series is increased. The relative order of performance remains constant though, with J_{GLS} still performing best. The increased power due to the longer time-series more clearly illustrates the effects of introducing skewness and kurtosis into the data. The more kurtosis is introduced the less power the test statistic has as is evidenced by the student t(3) and student t(5) graphs. Skewness has the most profound impact on power as is displayed in the reduction of the size-power curves of all statistics in the chi-squared graph.

When considering how J_{OLS}, J_{GLS}, J_{FM} and J_{GMM} are constructed, we want the test statistic that most often correctly rejects these statistics i.e. $H_3: \alpha = 0$. A holistic assessment would therefore advocate J_{GLS} as it correctly rejects **H_3** most often even though J_{OLS} has superior power properties.

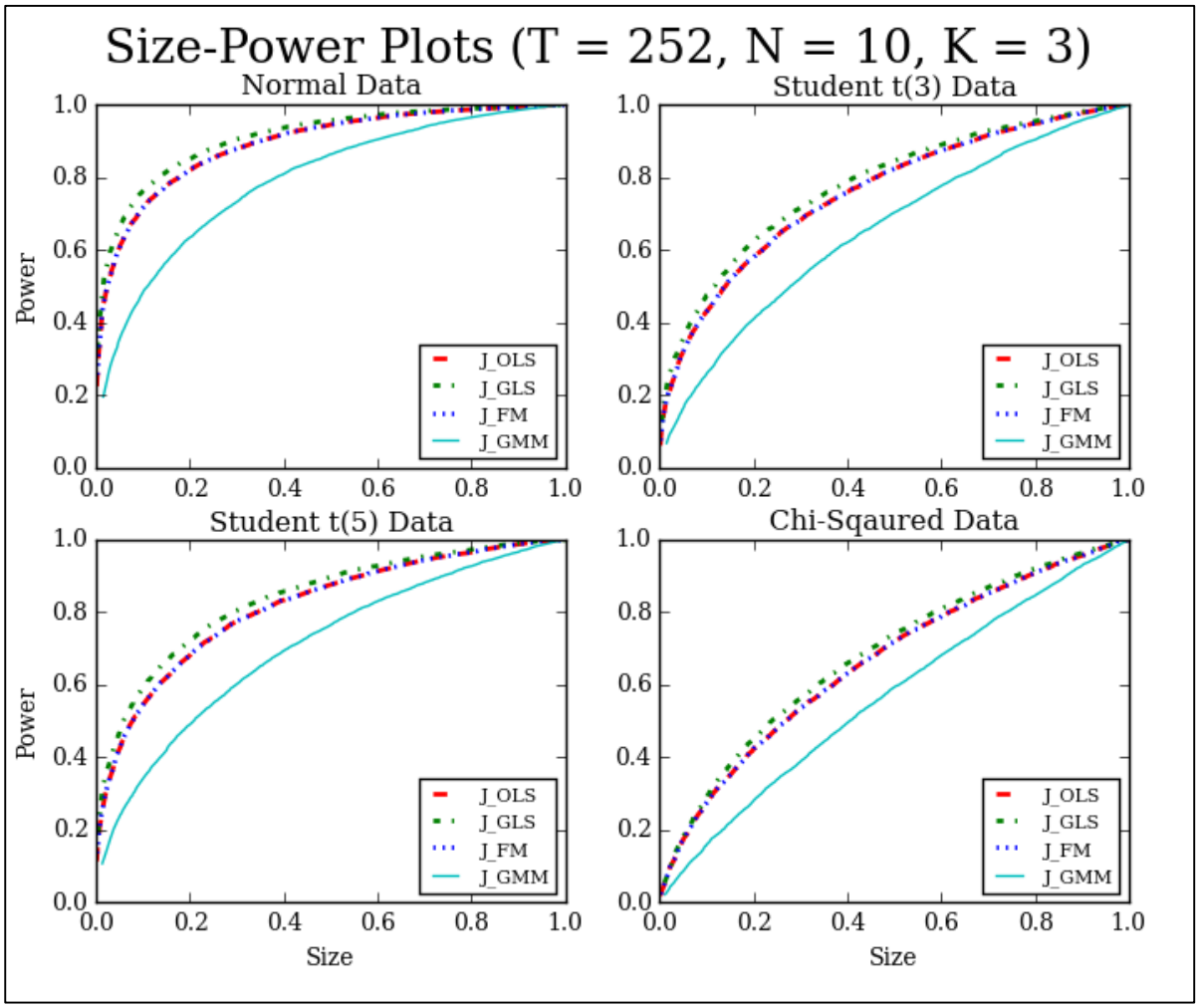


Figure 5.5 Size-power plots where T = 252

Finally a caveat, the errors are i.i.d. in the simulation and do not account for heteroskedasticity or autocorrelation. As autocorrelation is almost surely present in returns and J_{GMM} is designed to account for heteroskedasticity and autocorrelation, this might affect its performance. This is a possible area for future research.

6. Discussion and Conclusion

The linear CAPM is one of the most pervasive models in economic literature and is still widely used today. The principle idea behind the model is a linear relationship between expected returns and systematic risk. International and local empirical research has showed numerous contradictions of this relationship, casting serious doubt on the validity of the model. Implicit in the linear CAPM is the assumption of joint normal return distributions which Mandelbrot (1963) showed to be unrealistic more than 5 decades ago. The primary purpose of the study can be described as an attempt to address the failings of the linear CAPM, or linear asset pricing models for that matter, by relaxing the assumption of multivariate normality. This was done in two ways. The first is based on the introduction of higher moments into the CAPM, an approach that has shown to have merit in international literature. This study investigated the validity of the quadratic and cubic CAPM asset pricing models on the JSE. The second involved a Monte Carlo experiment to determine the impact of the assumption of normality on test statistics traditionally employed to assess the validity of asset pricing models. Non-normality of the sample data was established before addressing the primary objective of the study. Skewness and excess kurtosis was shown to be highly prevalent. This suggests that JSE listed shares are more prone to large, positive or negative, price movements than would be predicted by a normal distribution. Normality was strongly rejected on a univariate and a multivariate basis.

The study employed a wide range of analysis techniques which did not necessarily lead to the same conclusions. This leads to a need to review the results of this study in a holistic manner. In doing so we concluded that there is very little support for the hypothesis that higher moment CAPM models accurately characterise the relationship between risk and expected return. Both higher moment models were shown to have non-zero pricing errors on size sorted portfolios while pricing errors are insignificant on the value sorted portfolio set. The only definite conclusion is that results are dependent on the firm characteristic used in portfolio formation. This in turn would suggest that neither higher moment model is the true model. The true model should reject the null regardless of the data gathering process employed. Furthermore, the relative superior performance of the FF3F model casts doubt on Dittmar's (2002) assertion regarding the superiority of preference restricted non-linear pricing kernels over multifactor pricing kernels.

There is however moderately strong evidence that coskewness is priced and to a lesser degree that cokurtosis is priced. This indicates that investors do not only consider an asset's covariance with the market portfolio but also the asset's coskewness and cokurtosis. This is a fundamental shift from traditional mean-variance efficient portfolio selection. The literature reviewed suggests that the mean-variance optimal portfolios are, in all likelihood, sub-optimal in the presence of skewness (Vorkink and Mitton, 2007; Barberis and Huang, 2008; Harvey *et al.*,2010). Skewness is also proposed as an explanation for investor under-diversification. The idea is that portfolios that are mean-variance-skewness efficient would appear inefficient or under-diversified when viewed from a traditional mean-variance efficiency point of view.

Higher moments are found to be priced as on international stock exchanges and are in line with international literature. However the nature of the relationship of expected return with market beta,

coskewness beta and cokurtosis beta is in contrast with the international literature and even general asset pricing theory. The study confirms the Van Rensburg and Robertson (2003) finding of an inverse relationship between expected return and market beta. Contrary to Harvey and Siddique's (2000) hypothesis based on non-increasing absolute risk aversion, the results suggest a positive relationship between expected returns and coskewness. However this result is not conclusive as the size sorted portfolio set produces contradictory results. Contrary to Dittmar's (2002) hypothesis based on investor aversion to extreme outcomes, the results indicate a negative relationship between expected return and cokurtosis beta. The contradiction of Harvey and Siddique's (2000) and Dittmar's (2002) hypothesis casts serious doubt on the prudence of investors. These three results fundamentally undermine the linear, quadratic and cubic CAPMs since higher returns can be obtained by investing in less risky assets, where risk is defined as covariance, coskewness or cokurtosis with the market portfolio.

Higher moment models were not assessed in isolation but rather in comparison with the FF3F model. In most situations the FF3F model performed better than the higher moment models. The relative superior performance of the FF3F model casts doubt on Dittmar's (2002) assertion regarding the superiority of the preference restricted non-linear pricing kernels over multifactor pricing kernels. The estimation of the FF3F model by construction allows for an analysis of the size and value premiums which yielded interesting results. A periodic analysis revealed the size effect is concentrated in the January 1993 to December 1996 sub-period and a reversal of the size and value premia between January 2008 and December 2013. These findings appear to be robust as the cross-sectional regressions show that size and value factors are priced when estimated over January 1993 to December 2013 but not over the period January 1997 to December 2013.

In addressing the second part of the primary objective of the study, we set up a Monte Carlo experiment to determine the size and power characteristics of the test statistics used in the study. The simulation results when considered holistically indicate that OLS and GLS have superior size and power properties when evaluating non-normal data. The test of overidentifying restriction as formulated in Cochrane (2005) performs very poorly and as such should not be used on JSE data. It must be said though that the cross-sectional GMM test statistic could possibly perform better when autocorrelation and / or heteroskedasticity is present in the data.

The study is subject to a number of limitations that should be kept in mind when evaluating the results. Firstly, sample data is limited to the size sorted, value sorted and size-value sorted panel. The portfolios were selected due to their pervasive use in finance and economic literature to facilitate direct comparison. The study should be extended to other portfolios that have been shown to be correlated with expected return, most notably the momentum sorted portfolio set. The methodology of the paper is further limited in that asset returns are only assessed at a portfolio level, not at an individual security level. The extension to individual securities could materially impact all findings which makes it a crucial area for future research. Finally a caveat, in order to isolate the effects of skewness and kurtosis on test statistics, the Monte Carlo experiment was designed without incorporating autocorrelation and heteroskedasticity into errors. It has been proven that autocorrelation and heteroskedasticity can bias the outcome of a test statistic. Since it is very likely

that autocorrelation and heteroskedasticity are present within return data, certain estimation techniques have been designed to guard against their effects. J_{GMM} was in fact designed to account for heteroskedasticity and autocorrelation which could possibly have affected its performance in the Monte Carlo experiment. Expanding hte Monte Carlo experiment to incorporate not only skewness and kurtosis but heteroskedasticity and autocorrelationas well is another possible extension for future research.

References

Ang, A., & Chen, J. (2002). Asymmetric correlations of equity portfolios. *The Journal of Financial Economics, 63*, 443-494.

Ang, A., & Chen, J. (2007). CAPM over the long-run: 1926 – 2001. *The Journal of Emperical Finance, 14*, 1-40.

Auret, C. J., & Cline, R. (2011). Do the value, size and January effects exist on the JSE? *Investment Analysts Journal, 74*, 31-37.

Avramov, D., & Chordia, T. (2006). Asset pricing models and financial market anomalies. *Review of Financial Studies, 19*, 1001-1040.

Banz, R. W. (1981). The relationship between return and market value of common stocks. *The Journal of Financial Economics, 9*, 3-18.

Barberis, N. C., & Huang, M. (2008). Stocks as lotteries: The Implications of probability weighting for security prices. *Economic Review*, 2066-2100.

Barone-Adesi, G., Garliardini, P., & Urga, G. (2004). Testing asset pricing models with coskewness. *The Journal of Business & Economic Statistics, 22*, 474-485.

Basiewicz, P. G., & Auret, C. J. (2009). Another look at the cross-section of average returns on the JSE. *Investment Analysts Journal, 69*, 23-38.

Basiewicz, P. G., & Auret, C. J. (2010). Feasibility of the Fama and French three factor model in explaining returns on the JSE. *Investment Analysts Journal, 71*, 13 - 25.

Basu, S. (1977). Investment performance of common stocks in relation to their price-ratios: A test of the efficient market hypothesis. *The Journal of Finance, 32(3)*, 663-682.

Bhoraj, S., & Swaminathan, B. (2006). Macromomentum: Evidence of predictability in international equity markets. *The Journal of Business, 79*(1), 429–451.

Black, F. (1972). Capital market equilibrium with restricted borrowing. *The Journal of Business, 45*, 444-455.

Black, F., Jensen, M., & Scholes, M. (1972). The capital asset prcing model: Some empirical tests. *Studies in the theory of capital markets*, 79-121.

Blau, B. M., Masud, A., & Whitby, R. J. (2013). Does it pay to have fat tails? Examining kurtosis and the cross-section of stock returns. Retrieved from http://ssrn.com/abstract=2266042

Brown, P., Kleidon, P. W., & March, T. A. (1983). New evidence on size related anomalies in stock prices. *The Journal of Financial Economics, 12*, 3-12.

Cambell, J., Lo, A., & MacKinlay, C. (1996). *The econometrics of financial markets.* Princeton, NJ: Princeton University Press.

Carhart, M. M. (1997). On persistence in mutual fund performance. *The Journal of Finance, 52*, 57-82.

Chan, K. C., & Chen, N. (1988). An unconditional asset-pricing test and the role of firm size as an instrumental variable for risk. *The Journal of finance, 43*, 309-325.

Chan, L., Hamao, Y., & Lakonishok, J. (1991). Fundamentals and stock returns in Japan. *The Journal of Finance, 71*, 1739-1789.

Cochrane, J. (1999). New facts in finance. *The Journal of Economic Perspectives, 23*, 36-58.

Cochrane, J. (2005). *Asset pricing.* Princeton: Princeton University Press.

Conrad, J., Dittmar, R. F., & Ghysels, E. (2013). Ex ante skewness and expected stock returns. *The Journal of Finance, 68*, 85-124.

Daniel, K., & Titman, S. (1999). Market efficiency in an irrational world. *Financial Analysts Journal, 55*, 28-40.

Davidson, R., & MacKinnon, J. G. (1998). Graphical methods for investigating the size and power of hypothesis tests. *The Manchester School, 66*, 1-26.

De Moor, L., & Sercu, P. (2013). The smallest firm effect: an international study. *The Journal of International Money and Finance, 32*, 129-155.

Dimson, E., March, P., & Stauton, M. (2002). *Triumph of the optimists: 101 years of global investment returns.* New Jersey: Princeton University Press.

Dittmar, R. F. (2002). Nonlinear pricing kernels, kurtosis preference, and evidence from the cross section of equity returns. *The Journal of Finance, 57*, 369-403.

Doornik, J. A., & Hansen, H. (2008). An omnibus test for univariate and multivariate normality. *Oxford Bulletin of Economics and Statistics, 70*, 927-939.

Fama, E. F. (1965). The behaviour of stock market prices. *The Journal of Business, 38*, 34-105.

Fama, E. F. (1970). Efficient capital markets: a review of theory and empirical work. *The Journal of Finance, 25*, 383-417.

Fama, E. F., & French, K. R. (1992). The cross-section of expected stock returns. *The Journal of Finance, 47*(2), 430-465.

Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *The Journal of Finance, 50*(1), 1-56.

Fama, E. F., & French, K. R. (1996). Multifactor explanations of asset-pricing anomalies. *The Journal of Finance, 47*, 426-465.

Fama, E. F., & French, K. R. (2006). The value premium and the CAPM. *The Journal of Finance, 61*, 2163–2185.

Fama, E. F., & French, K. R. (2008). Dissecting anomalies. *The Journal of Finance, 63*, 1653-1678.

Fama, E. F., & French, K. R. (2012). Size, value, and momentum in international stock returns. *The Journal of Financial Economics, 105*, 457-472.

Fama, J., & McBeth, J. (1973). Risk, return and equilibrium: Empirical tests. *Journal of Political Economy, 81*, 607-636.

Fang, H., & Lai, T. (1997). Co-kutosis and capital asset pricing. *The Financial Review, 32*, 293-307.

Ferguson, M. F., & Shockley, L. R. (2003). Equilibrium anomalies. *The Journal of Finance, 58*, 2549-2580.

Friend, I., & Blume, M. (1970). Measurement of portfolio performance under uncertainty. *The American Economic Review, 60*(4), 561-575.

Gompers, P. A., & Metrick, A. (2001). Institutional investors and equity prices. *The Quarterly Journal of Economics, 116*(1), 229-259.

Hahn, J., & Lee, H. (2006). Yield spreads as alternative risk factors for size and book-to-market. *The Journal of Financial and Quantitative Analysis, 41*, 245-269.

Hansen, L. (1982). Large smpale properties of generalized method of moments estimators. *Econometrica, 50*, 1029-1054.

Harvey, C. R., Liechty, J. C., Liechty, M. W., & Muller, P. (2010). Portfolio selection with higher moments. *Quantitative Finance, 10*, 469-485.

Harvey, C., & Siddique, A. (2000). Conditional skewness in asset pricing tests. *The Journal of Finance, 55*, 1263-1295.

Harvey, C., & Zhou, G. (1993). Conditional skewness in asset pricing. *The Journal of Empirical Finance, 1*, 107-131.

Henze, N., & Zirkler, B. (1990). A class of invariant consistent tests for multivariate normality. *Communications in Statistics - Theory and Methods, 19*, 3595-3617.

Hodgson, D., Linton, O., & Vorkink, K. (2002). Testing the capital asset pricing model efficiently under elliptical symmetry: A semiparametric approach. *The Journal of Applied Econometrics, 17*(6), 617-639.

Hoffman, A. J. (2012). Stock return anomalies: Evidence from the Johannesburg Stock Exchange. *Investment Analysts Journal, 75*, 21-41.

Huang, P., & Hueng, J. C. (2008). Conditional risk–return relationship in a time-varying beta model. *The Journal of Quantitative Finance, 8*(4), 381-390.

Hwang, S., & Satchell, S. (1999). Modelling emerging market risk premia using higher moments. *International Journal of Finance & Economics, 4*, 271-296.

Jagannathan, R., & Wang, Z. (1996). The conditional CAPM and the cross-section of expected returns. *The Jounal of Finance, 61*, 3-53.

Janse van Rensburg, S., Friskin, D., & Sharp, G. D. (2009). A distributional comparison of size-based portfolios on the JSE. *Investment Analysts Journal*(69), 17-22.

Jarque, C. M., & Bera, A. K. (1987). A test for normality of observations and regression residuals. *International Statistical Review, 55*(2), 163-172.

Jegadeesh, N., & Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal Finance, 48(1)*, 65-91.

Jensen, M. C. (1968). The performance of mutual funbs in the period 1945-1964. *The Journal of Finance, 23*, 389-416.

Keim, D. B. (1983). Size related anomalies and stock return seasonality: Empirical evidence. *The Journal of Financial Economics, 12*, 12-32.

Kimball, M. (1990). Precautionary saving in the small and in the large. *Econometrica, 58*, 53-73.

Kraus, A., & Litzenberger, R. (1976). Skewness preference and the valuation of risky assets. *The Journal of Finance, 21*, 1085-1094.

Lakonishok, J., Shleifer, A., & Vishny, R. (1994). Contrarion investment, extrapolation, and risk. *The Journal of Finance*, 1541-1578.

Lettau, M., & Ludvigson, S. (2001). Resurrecting the (C)CAPM: A cross-sectional test when risk premia are time-varying. *The Journal of Political Economy, 109*, 1238-1287.

Lewellen, J. (2002). Momentum and autocorrelation in stock returns. *Review of Financial Studies, 15*, 533-563.

Lewellen, J., & Nagel, S. (2006). The conditional CAPM does not explain asset pricing anomalies. *The Journal of Financial Economics, 82*, 289-314.

Lim, K. (1989). A new test of the three-moment capital asset pricing model. *The Journal of Financial and Quantitative Analysis, 24*(2), 205-216.

Lintner, J. (1965). The valuation of risk assets and the selection of risky investments in stock portfolios and capital assets. *Review of Economics & Statistics, 47*, 13-37.

Liu, W. (2006). A liquidity-augmented capital asset pricing model. *The Journal of Financial Economics, 82*, 631–671.

Lo, A., & MacKinlay, C. (1990). Data-snooping biases in tests of financial asset pricing models. *The Review of Financial Studies, 3*, 431-467.

Lustig, H., & Van Nieuwerburgh, S. (2003). Housing collateral, consumption insurance, and risk permia. *Working paper, University of Chicago*.

Mandelbrot, B. (1963). The variation of certain speculative prices. *The Journal of Business, 36*, 394-419.

Mangani, R. (2007). Distributional properties of JSE Returns. *Investment Analysts Journal, 66*, 57-72.

Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika, 57*, 519-530.

Markowitz, H. (1952). Portfolio selection. *The Journal of Finance, 7*, 77-91.

Markowitz, H. (1959). *Portfolio selection: Efficient diversification of investments.* New York: Wiley.

Mayers, D. (1972). Nonmarketable assets and capital market equilibrium under uncertainty. In M. C. Jensen, *Studies in the Theory of Capital Markets* (pp. 223-248). New York: Praeger.

McClelland, S. E., Auret, C. J., & Wright, T. K. (2014). Thin-trading and beta estimation: Results from a simulated environment. *The Journal for Studies in Economics and Econometrics, 38*, 19-32.

Merton, R. C. (1973). Theory of rational option pricing. *The Bell Journal of Economics and Management Science, 4*(1), 141-183.

Mitton, T., & Vorkink, K. (2007). Equilibrium underdiversification and the preference for skewness. *Review of Financial Studies, 20*, 1255-1288.

Mossin, J. (1966). Equilibrium in the capital asset market. *Econometrica, 34*, 768-783.

Newey, W. K., & West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica, 55*(3), 703-708.

Nijman, T., Swinkels, L., & Verbeek, M. (2004). Do countries or industries explain momentum in Europe. *The Journal of Empirical Finance, 11*, 461-481.

Page, D. (2012). Reviving beta? Another look at the cross-section of average share returns on the JSE. *Masters dissertation, The University of the Witwatersrand*.

Petkova, P. (2006). Do the Fama–French factors proxy for innovations in predictive variables. *The Journal of Finance, 61*, 581-612.

Petkova, P., & Zhang, L. (2005). Is value is riskier than growth? *The Journal of Financial Economics, 78*, 187-202.

Pettengill, G. N., Sundaram, S., & Mathur, I. (1995). The conditional relation between beta and returns. *The Journal of Financial and Quantitative Analysis, 30*, 101-116.

Ramsey, J. B. (1969). Test for specification errors in classical linear least squares regression analysis. *The Journal of the Royal Statistical Society, Series B, 31*, 350-371.

Reinganum, R. M. (1981). Misspecification in capital asset pricing: empirical anomalies based on earnings yields and market values. *The Journal of Finacial Economics, 9*, 19-46.

Robins, E. M., Sandler, M., & Durand, F. (1999). Inter-relationships between the January effect, market capitalisation and value investment strategies on the JSE. *The Investment Analyst Journal, 50*, 53-64.

Roll, R. (1977). A critique of the asset pricing theory's tests part1: On past and potential testability of the theory. *The Journal of Financial Economics, 4*(2), 129-176.

Roll, R. (1983). The turn-of-the-year effect and the return premia of small firms. *The Journal of Portfolio Managment*, 18-28.

Rosenberg, B., Reid, K., & Lanstein, R. (1985). Persuasive evidence of market inefficiency. *The Journal of Portfolio Management, 11*, 9-17.

Ross, S. A. (1976). The arbitrage pricing theory of capital asset pricing. *The Journal of Economic Theory, 13*(3), 341-360.

Royston, J. P. (1983). Some techniques for assessing multivariate normality based on the Shapiro-Wilk. *The Journal of Applied Statistics, 32*, 121-133.

Rozeff, M. S., & Kinney, W. R. (1976). Capital market seasonality: The case of stock returns. *The Journal of Financial Economics, 3*, 379-402.

Rubenstein, M. (1973). The fundamental theorem of parameter preference security security valuation. *The Journal of Financial and Qualitative Analysis, 8*, 61-69.

Sears, R. S., & Wei, K. C. (1985). Asset pricing, higher moments and the market risk premium: A note. *The Journal of Finance, 11*(4), 1251-1253.

Shadwick, W., & Keeting, C. (2002). A universal performance measure. *The Journal of Performance Measurement*, 59-84.

Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality. *Biometrika, 52*(3), 591-611.

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance, 19*, 425–442.

Siriwardane, E. (2013). Notes on the econometrics of empirical asset pricing. *Working Paper, NYU Stern School of Business*.

Strugnell, D., Gilbert, E., & Kruger, R. (2011). Beta, size and value effects on the JSE, 1994-2007. *Investment Analysts Journal, 74*, 1-17.

Tobin, J. (1958). Liquidity preference as behavior towards risk. *The Review of Economic Studies, 25*, 65-86.

Treynor, J. L. (1963). *Implications for the theory of finance.* Unpublished manuscript.

Treynor, J. L. (1965). How to rate the managment of investment funds. *Harvard Business Review*, 63-75.

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *The Journal of Risk and Uncertainty, 5*, 297-323.

Van Rensburg, P., & Robertson, M. (2003). Size, price-to-earnings and beta on the JSE securities excahnge. *Investment Analyst Journal*(58), 7-16.

Van Rensburg, P., & Slanley, K. B. (1997). Market segmentation on the Johannesburg Stock Exchange. *The Journal for Studies in Economics and Econometrics, 23*(3), 1-23.

Vorkink, K. (2003). Return distributions and improved tests of asset pricing models. *The Review of Financial Studies, 16*, 845-874.

Wang, K. (2002). Asset pricing with conditioning information: A new test. *The Journal of Finance, 58*, 161-196.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica, 48*(4), 817-38.

Williams, J. (1977). Capital asset prices with heterogeneous beliefs. *The Journal of financial economics, 5*, 219-239.

Xiong, J. X., & Idozorek, T. (2011). The impact of skewness and fat tails on the asset allocation decision. *The Financial Analyst Journal, 67*, 23-25.

Zhang, L. (2005). The value premium. *The Journal of Finance, 60*, 67-101.

Appendix A: Portfolio and Factor Formation Automation

This appendix details the technology and processes utilised to automate portfolio construction and factor formation. The overall portfolio formation process is illustrated in figure B1 below.

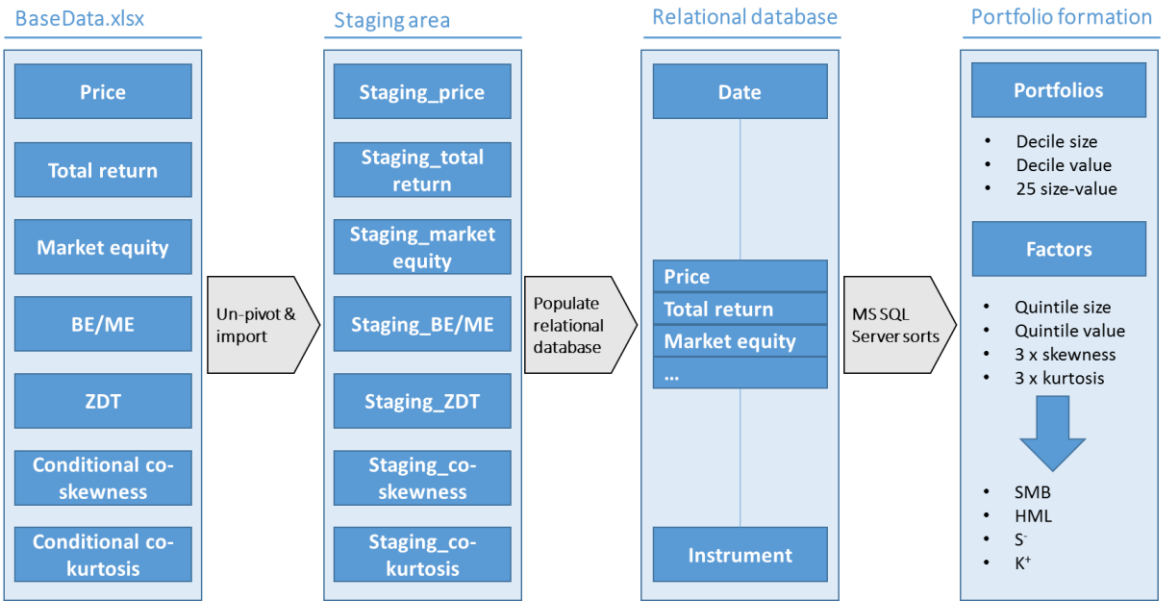
Database Creation

Base data was retrieved from the Findata@Wits database for the period January 1992 to December 2013 in Microsoft (MS henceforth) Excel format. The following measures were obtained for all JSE listed securities:

- Closing price
- Total return[1]
- Market equity
- Book equity (BE) / market equity (ME)
- Zero days trading (ZDT)

Conditional co-skewness and conditional co-kurtosis were calculated for the period January 1997 and December 2013 using MS Visual Basic for Applications (VBA). Each measure populates a worksheet within a single MS Excel file.

Figure B1: Database creation process



The data within each worksheet are then un-pivoted and imported using MS SQL Server Integration Services in order to create a relational database. The base data in each worksheet follows the format illustrated in table B1:

Table B1: Base data format

[1] Please refer to section 3 for calculation

Instrument	Date_1	Date_2	…	Date_N
DBR	Value_1	Value_2	…	Value_N
ACC	Value_1	Value_2	…	Value_N
…	…	…	…	…

Data are un-pivoted into panel data format to allow for the creation of a relational database that facilitates the automation of portfolio sorts as illustrated in table B2.

Table B2: Panel data format

Instrument	Date	Value
DBR	Date_1	Value_1
DBR	Date_2	Value_2
…	…	…
DBR	Date_N	Value_N
ACC	Date_1	Value_1
ACC	Date_2	Value_2
…	…	…
ACC	Date_N	Value_N
…	…	…

Panel data are imported with dimensions date and instrument populating price, total return, market equity, BE/ME, ZDT, conditional co-skewness and conditional co-kurtosis. Please refer figure B2

Figure B2: MS SQL Server 2012 code to populate dimensions

```
-- Update date dimension[1]

INSERT INTO DimDate (YearMonth, [Year], [YYYYMM], [Month])
SELECT DISTINCT [YearMonth], DATEPART(yy, CONVERT(date, [YearMonth])), CONVERT(char(6),
CONVERT(date, [YearMonth]), 112), DATEPART(mm, CONVERT(date, YearMonth))
FROM (
SELECT COALESCE(B.YearMonth, MC.[YearMonth], P.[YearMonth], TR.[YearMonth], Z.[YearMonth],
S.[YearMonth], K.[YearMonth], LNMC.[YearMonth]) AS YearMonth
FROM Staging_BM AS B
FULL OUTER JOIN Staging_MarketCap AS MC
ON B.[First True Code] = MC.[First True Code]
AND B.[YearMonth] = MC.[YearMonth]
FULL OUTER JOIN Staging_Price AS P
ON B.[First True Code] = P.[First True Code]
AND B.[YearMonth] = P.[YearMonth]
FULL OUTER JOIN Staging_TotalReturns AS TR
ON B.[First True Code] = TR.[First True Code]
AND B.[YearMonth] = TR.[YearMonth]
FULL OUTER JOIN Staging_ZDT AS Z
ON B.[First True Code] = Z.[First True Code]
AND B.[YearMonth] = Z.[YearMonth]
FULL OUTER JOIN Staging_S AS S
ON B.[First True Code] = S.[First True Code]
AND B.[YearMonth] = S.[YearMonth]
```

[1] -- Refers to comments

```
FULL OUTER JOIN Staging_K AS K
ON B.[First True Code] = K.[First True Code]
AND B.[YearMonth] = K.[YearMonth]
FULL OUTER JOIN Staging_LNMC AS LNMC
ON B.[First True Code] = LNMC.[First True Code]
AND B.[YearMonth] = LNMC.[YearMonth]
WHERE   COALESCE(B.YearMonth,   MC.[YearMonth],   P.[YearMonth],   TR.[YearMonth],   Z.[YearMonth],
S.[YearMonth], K.[YearMonth], LNMC.[YearMonth]) IS NOT NULL
) AS x

-- Update instrument dimension

INSERT INTO DimInstrument (InstrumentCode)
SELECT DISTINCT InstrumentCode
FROM (
SELECT COALESCE( B.[First True Code], MC.[First True Code], P.[First True Code], TR.[First True
Code], Z.[First True Code], S.[First True Code], K.[First True Code], LNMC.[First True Code]) AS
InstrumentCode
FROM Staging_BM AS B
FULL OUTER JOIN Staging_MarketCap AS MC
ON B.[First True Code] = MC.[First True Code]
AND B.[YearMonth] = MC.[YearMonth]
FULL OUTER JOIN Staging_Price AS P
ON B.[First True Code] = P.[First True Code]
AND B.[YearMonth] = P.[YearMonth]
FULL OUTER JOIN Staging_TotalReturns AS TR
ON B.[First True Code] = TR.[First True Code]
AND B.[YearMonth] = TR.[YearMonth]
FULL OUTER JOIN Staging_ZDT AS Z
ON B.[First True Code] = Z.[First True Code]
AND B.[YearMonth] = Z.[YearMonth]
FULL OUTER JOIN Staging_S AS S
ON B.[First True Code] = S.[First True Code]
AND B.[YearMonth] = S.[YearMonth]
FULL OUTER JOIN Staging_K AS K
ON B.[First True Code] = K.[First True Code]
AND B.[YearMonth] = K.[YearMonth]
FULL OUTER JOIN Staging_LNMC AS LNMC
ON B.[First True Code] = LNMC.[First True Code]
AND B.[YearMonth] = LNMC.[YearMonth] ) AS x
```

Figure B3 outlines the creation and population of a table (FactJSE) from which we will be to complete portfolios.

Figure B3: MS SQL Server 2012 code to create and populate FactJSE

```
-- Populate FactJSE[1]

INSERT INTO FactJSE
( InstrumentID, DateID, BM, MarketCap, Price, TotalReturns, ZDT, S, K, LNMC )
SELECT i.InstrumentID, d.DateID, ds.BM, ds.MarketCap, ds.Price, ds.TotalReturns, ds.ZDT, ds.S,
ds.K, ds.LNMC
```

[1] -- Refers to comments

```
FROM (
SELECT COALESCE( B.[First True Code], MC.[First True Code], P.[First True Code], TR.[First True
Code], Z.[First True Code], S.[First True Code], K.[First True Code], LNMC.[First True Code]) AS
InstrumentCode,
CASE WHEN ISNUMERIC(B.Value) = 1 THEN CONVERT(float, REPLACE (B.Value, ',', '.')) ELSE NULL END AS
BM,
IIF(ISNUMERIC(P.Value)=1,CONVERT(float,REPLACE (P.Value, ',', '.')), NULL) AS Price,
IIF(ISNUMERIC(TR.Value)=1,CONVERT(float,REPLACE (TR.Value, ',', '.')), NULL) AS TotalReturns,
COALESCE(B.YearMonth, MC.[YearMonth], P.[YearMonth], TR.[YearMonth], Z.[YearMonth], S.[YearMonth],
K.[YearMonth], LNMC.[YearMonth]) AS YearMonth,
IIF(ISNUMERIC(MC.Value)=1,CONVERT(float,REPLACE (MC.Value, ',', '.')), NULL) AS MarketCap,
IIF(ISNUMERIC(Z.Value)=1,CONVERT(float,REPLACE (Z.Value, ',', '.')), NULL) AS ZDT,
IIF(ISNUMERIC(S.Value)=1,CONVERT(float,REPLACE (S.Value, ',', '.')), NULL) AS S,
IIF(ISNUMERIC(K.Value)=1,CONVERT(float,REPLACE (K.Value, ',', '.')), NULL) AS K,
IIF(ISNUMERIC(LNMC.Value)=1,CONVERT(float,REPLACE (LNMC.Value, ',', '.')), NULL) AS LNMC
FROM Staging_BM AS B
FULL OUTER JOIN Staging_MarketCap AS MC
ON B.[First True Code] = MC.[First True Code]
AND B.[YearMonth] = MC.[YearMonth]
FULL OUTER JOIN Staging_Price AS P
ON B.[First True Code] = P.[First True Code]
AND B.[YearMonth] = P.[YearMonth]
FULL OUTER JOIN Staging_TotalReturns AS TR
ON B.[First True Code] = TR.[First True Code]
AND B.[YearMonth] = TR.[YearMonth]
FULL OUTER JOIN Staging_ZDT AS Z
ON B.[First True Code] = Z.[First True Code]
AND B.[YearMonth] = Z.[YearMonth]
FULL OUTER JOIN Staging_S AS S
ON B.[First True Code] = S.[First True Code]
AND B.[YearMonth] = S.[YearMonth]
FULL OUTER JOIN Staging_K AS K
ON B.[First True Code] = K.[First True Code]
AND B.[YearMonth] = K.[YearMonth]
FULL OUTER JOIN Staging_LNMC AS LNMC
ON B.[First True Code] = LNMC.[First True Code]
AND B.[YearMonth] = LNMC.[YearMonth]
WHERE   COALESCE(B.YearMonth,   MC.[YearMonth],   P.[YearMonth],   TR.[YearMonth],   Z.[YearMonth],
S.[YearMonth], K.[YearMonth], LNMC.[YearMonth]) IS NOT NULL
) AS DS
JOIN DimDate d
ON ds.YearMonth = d.YearMonth
JOIN DimInstrument I
ON ds.InstrumentCode = i.InstrumentCode
WHERE ds.MarketCap IS NOT NULL
AND ds.Price IS NOT NULL
AND ds.TotalReturns IS NOT NULL
AND ds.ZDT IS NOT NULL
```

Data sorting

Once the fact table has been created the liquidity filter is applied before data are sorted into size, value, size-value, conditional co-skewness and conditional co-kurtosis portfolios. Please refer to figure B4 for detailed description of sorting procedures.

Figure B4: MS SQL Server 2012 code to sort data into portfolios

```
-- Apply liquidity filter[1]

UPDATE F
SET LiquidityFilter = 0
FROM vwFactJSE AS f

UPDATE f
SET LiquidityFilter = 1
FROM vwFactJSE AS f
JOIN (
SELECT [Year], InstrumentCode
FROM vwFactJSE
GROUP BY [Year], InstrumentCode
HAVING COUNT(*) = 12
AND SUM(ZDT) < 150 ) AS x
ON f.[Year] = x.[year]
AND f.InstrumentCode = x.InstrumentCode

-- Calculate 12 month averages of measures for sorting

INSERT INTO FactJSEYear
( InstrumentCode, [Year], LiquidityFilter, Price, TotalReturns, BM, MarketCap, S, K, LNMC )
SELECT x.InstrumentCode, x.[Year], x.LiquidityFilter, x.Price, x.TotalReturns,
        x.BM, x.MarketCap, y.S, y.K,x.LNMC
FROM (
        SELECT InstrumentCode, [Year], LiquidityFilter, AVG(Price) as Price, AVG(TotalReturns) AS
TotalReturns,
        AVG(BM) AS BM, AVG(MarketCap) AS MarketCap, AVG(LNMC) AS LNMC
        FROM vwFactJSE x
        GROUP BY InstrumentCode, [Year], LiquidityFilter
        ) AS x
JOIN (
        SELECT InstrumentCode, [Year], S, K
        FROM vwFactJSE
        WHERE [Month] = 1
        ) AS y
        ON x.InstrumentCode = y.InstrumentCode
        AND x.[Year] = y.[Year];
GO

-- Portfolio sorts

DECLARE @i int = 1992;

WHILE @i <= 2014
BEGIN
-- Quintile and decile portfolio sorts on average market capitalisation

        UPDATE f
        SET SizeFilter =
                CASE
                        WHEN x.QTile = 1 THEN 'S'
```

[1] -- Refers to comments

```
                        WHEN x.QTile = 2 THEN '2'
                        WHEN x.QTile = 3 THEN '3'
                        WHEN x.QTile = 4 THEN '4'
                        ELSE 'B'
                END
        FROM FactJSEYear AS f
        JOIN (
                SELECT InstrumentCode,
                        NTILE(5) OVER (ORDER BY MarketCap) AS QTile
                FROM FactJSEYear
                WHERE LiquidityFilter = 1
                AND ISNULL(MarketCap, 0) > 0
                AND [Year] = @i-1
                ) AS x
        ON f.InstrumentCode = x.InstrumentCode
        WHERE ISNULL(f.MarketCap, 0) > 0
        AND f.[Year] = @i;


        UPDATE f
        SET SizeFilter10 =
                CASE
                        WHEN x.QTile = 1 THEN 'S'
                        WHEN x.QTile = 2 THEN '2'
                        WHEN x.QTile = 3 THEN '3'
                        WHEN x.QTile = 4 THEN '4'
                        WHEN x.QTile = 5 THEN '5'
                        WHEN x.QTile = 6 THEN '6'
                        WHEN x.QTile = 7 THEN '7'
                        WHEN x.QTile = 8 THEN '8'
                        WHEN x.QTile = 9 THEN '9'
                        ELSE 'B'
                END
        FROM FactJSEYear AS f
        JOIN (
                SELECT InstrumentCode,
                        NTILE(10) OVER (ORDER BY MarketCap) AS QTile
                FROM FactJSEYear
                WHERE LiquidityFilter = 1
                AND ISNULL(MarketCap, 0) > 0
                AND [Year] = @i-1
                ) AS x
        ON f.InstrumentCode = x.InstrumentCode
        WHERE ISNULL(f.MarketCap, 0) > 0
        AND f.[Year] = @i;

-- Quintile and decile portfolio sorts on average BE/ME

       UPDATE f
        SET ValueFilter =
                CASE
                        WHEN x.QTile = 1 THEN 'L'
                        WHEN x.QTile = 2 THEN '2'
                        WHEN x.QTile = 3 THEN '3'
                        WHEN x.QTile = 4 THEN '4'
                        ELSE 'H'
                END
        FROM FactJSEYear AS f
        JOIN (
```

```
                SELECT InstrumentCode,
                        NTILE(5) OVER (ORDER BY BM) AS QTile
                FROM FactJSEYear
                WHERE LiquidityFilter = 1
                AND ISNULL(BM, 0) > 0
                AND [Year] = @i-1
                ) AS x
        ON f.InstrumentCode = x.InstrumentCode
        WHERE ISNULL(f.BM, 0) > 0
        AND f.[Year] = @i;

        UPDATE f
        SET ValueFilter10 =
                CASE
                        WHEN x.QTile = 1 THEN 'L'
                        WHEN x.QTile = 2 THEN '2'
                        WHEN x.QTile = 3 THEN '3'
                        WHEN x.QTile = 4 THEN '4'
                        WHEN x.QTile = 5 THEN '5'
                        WHEN x.QTile = 6 THEN '6'
                        WHEN x.QTile = 7 THEN '7'
                        WHEN x.QTile = 8 THEN '8'
                        WHEN x.QTile = 9 THEN '9'
                        ELSE 'H'
                END
        FROM FactJSEYear AS f
        JOIN (
                SELECT InstrumentCode,
                        NTILE(10) OVER (ORDER BY BM) AS QTile
                FROM FactJSEYear
                WHERE LiquidityFilter = 1
                AND ISNULL(BM, 0) > 0
                AND [Year] = @i-1
                ) AS x
        ON f.InstrumentCode = x.InstrumentCode
        WHERE ISNULL(f.BM, 0) > 0
        AND f.[Year] = @i;

-- 5*5 Size-Value sorts to create 25 size-value portfolios

        WITH ctX AS
        (
                SELECT x.InstrumentCode, x.SizeFilter, y.BM, y.MarketCap
                FROM  FactJSEYear x
                JOIN  FactJSEYear y
                ON x.InstrumentCode = y.InstrumentCode
                WHERE x.[Year] = @i
                AND y.[Year] = @i-1
                AND y.LiquidityFilter = 1
                AND ISNULL(y.BM, 0.0) > 0
        )
        , ctY AS
        (
                SELECT
                        'S' AS SizeFilter,
                        InstrumentCode,
                        NTILE(5) OVER (ORDER BY BM) AS QTile
                FROM ctX
                WHERE SizeFilter = 'S'
```

```
                UNION
                SELECT
                        '2' AS SizeFilter,
                        InstrumentCode,
                        NTILE(5) OVER (ORDER BY BM) AS QTile
                FROM ctX
                WHERE SizeFilter = '2'
                UNION
                SELECT
                        '3',
                        InstrumentCode,
                        NTILE(5) OVER (ORDER BY BM) AS QTile
                FROM ctX
                WHERE SizeFilter = '3'
                UNION
                SELECT
                        '4',
                        InstrumentCode,
                        NTILE(5) OVER (ORDER BY BM) AS QTile
                FROM ctX
                WHERE SizeFilter = '4'
                UNION
                SELECT
                        'B',
                        InstrumentCode,
                        NTILE(5) OVER (ORDER BY BM) AS QTile
                FROM ctX
                WHERE SizeFilter = 'B'
        )
        UPDATE f
        SeT SizeValueFilter =
                CASE
                        WHEN ctY.QTile = 1 THEN 'L'
                        WHEN ctY.QTile = 2 THEN '2'
                        WHEN ctY.QTile = 3 THEN '3'
                        WHEN ctY.QTile = 4 THEN '4'
                        ELSE 'H'
                END
        FROM FactJSEYear f
        JOIN ctY
        ON f.InstrumentCode = ctY.InstrumentCode
        AND f.SizeFilter = ctY.SizeFilter
        AND f.[Year] = @i;

        SET @i = @i + 1
END
GO

-- Skewness factor portfolio creation

       UPDATE f
        SET SFilter =
                CASE
                        WHEN x.QTile = 1 THEN '-'
                        WHEN x.QTile = 2 THEN 'M'
                        ELSE '+'
                END
        FROM FactJSEYear AS f
        JOIN (
```

```
                SELECT InstrumentCode,
                        NTILE(3) OVER (ORDER BY S) AS QTile
                FROM FactJSEYear
                WHERE LiquidityFilter = 1
                AND S IS NOT NULL
                AND [Year] = @i
                ) AS x
        ON f.InstrumentCode = x.InstrumentCode
        AND f.LiquidityFilter = 1
        AND S IS NOT NULL
        AND f.[Year] = @i;

-- Kurtosis factor portfolio creation

UPDATE f
        SET KFilter = QTile
        FROM FactJSEYear AS f
        JOIN (
                SELECT InstrumentCode,
                        NTILE(3) OVER (ORDER BY K) AS QTile
                FROM FactJSEYear
                WHERE LiquidityFilter = 1
                AND K IS NOT NULL
                AND [Year] = @i
                ) AS x
        ON f.InstrumentCode = x.InstrumentCode
        AND f.LiquidityFilter = 1
        AND K IS NOT NULL
        AND f.[Year] = @i;
END
```

Portfolio formation

The end result of all the MS SQL Server 2012 code above is a database that is compatible with MS Excel. MS Excel PowerPivot is used to import the FactJSE table from the created database as a MS Excel data model. The data model is then converted into a pivot table. Finally, the pivot table can be manipulated to show portfolio price, total return, market equity, BE/ME, ZDT, conditional coskewness and conditional cokurtosis for differing periods. The pivot table is manually manipulated to create portfolios.

Appendix B: Higher Moment Automation using Visual Basic for Applications

This appendix details the Visual Basic for Applications (VBA) code utilised to calculate unconditional coskewness, unconditional cokurtosis, conditional coskewness and conditional cokurtosis. All functions have two input variables, *Share_Ret* and *Market_Ret*. *Share_Ret* is a Microsoft (MS) Excel range object of length T that corresponds to R^e, a $1 \times T$ vector of asset or portfolio excess returns over the risk free rate. *Market_Ret* is a MS Excel range object of length T that corresponds to R^e, a $1 \times T$ vector of the market portfolio excess returns over the risk free rate.

Function *UCoSkew* calculates unconditional coskewness, β_i^{m2}, as per Equation 2.19:

$$\beta_i^{m2} = \frac{Cov\left[R^{ei}(Re^m)^2\right]}{E\left[\left(R^{em} - E(R^{em})\right)^3\right]}$$

Where β_i^{m2} is defined in Equation 2.20:

$$R^i - R^f = \alpha^i + \beta_i^m\left(R^m - R^f\right) + \beta_i^{m2}\left(R^m - R^f\right)^2 + \varepsilon^i$$

Function *UCoSkewT* calculated the t-statistic of β_i^{m2}.

Figure C1: VBA code to calculate unconditional coskewness

```
Function UCoSkew(ByRef Share_Ret As Range, ByRef Market_Ret As Range)

    T = Share_Ret.Count
    ReDim Factor_Mat(1 To T, 1 To 2) As Double

    For i = 1 To T
        Factor_Mat(i, 1) = Market_Ret(i)
        Factor_Mat(i, 2) = Market_Ret(i) ^ 2
    Next i

    x = WorksheetFunction.LinEst(Share_Ret, Factor_Mat, True, True)

    UCoSkew = x(1, 1)

End Function

Function UCoSkewT(ByRef Share_Ret As Range, ByRef Market_Ret As Range)

    T = Share_Ret.Count
    ReDim Factor_Mat(1 To T, 1 To 2) As Double

    For i = 1 To T
        Factor_Mat(i, 1) = Market_Ret(i)
        Factor_Mat(i, 2) = Market_Ret(i) ^ 2
    Next i

    x = WorksheetFunction.LinEst(Share_Ret, Factor_Mat, True, True)

    UCoSkewT = x(1, 1) / x(2, 1)

End Function
```

Function *UCoKurt* calculates unconditional cokurtosis, β_i^{m3}, as per Equation 2.24:

$$\beta_i^{m3} = \frac{Cov\left[R^{ei}(R^{em})^3\right]}{E\left[\left(R^{em} - E(R^{em})\right)^4\right]}$$

Where β_i^{m3} is defined in Equation 2.23:

Function *UCoKurtT* calculated the t-statistic ofβ_i^{m3}.

Figure C2: VBA code to calculate unconditional cokurtosis

```
Function UCoKurt(ByRef Share_Ret As Range, ByRef Market_Ret As Range)

    T = Share_Ret.Count
    ReDim Factor_Mat(1 To T, 1 To 3) As Double


    For i = 1 To T
        Factor_Mat(i, 1) = Market_Ret(i)
        Factor_Mat(i, 2) = Market_Ret(i) ^ 2
        Factor_Mat(i, 3) = Market_Ret(i) ^ 3
    Next i

    x = WorksheetFunction.LinEst(Share_Ret, Factor_Mat, True, True)

    UCoKurt = x(1, 1)

End Function

Function UCoKurtT(ByRef Share_Ret As Range, ByRef Market_Ret As Range)

    T = Share_Ret.Count
    ReDim Factor_Mat(1 To T, 1 To 3) As Double


    For i = 1 To T
        Factor_Mat(i, 1) = Market_Ret(i)
        Factor_Mat(i, 2) = Market_Ret(i) ^ 2
        Factor_Mat(i, 3) = Market_Ret(i) ^ 3
    Next i

    x = WorksheetFunction.LinEst(Share_Ret, Factor_Mat, True, True)

    UCoKurtT = x(1, 1) / x(2, 1)


End Function
```

Function *CoSkew* calculates conditional coskewness, $\beta_{i,t}^{sks}$, as per Equation 2.22:

$$\beta_{i,t}^{sks} = \frac{E\left(\varepsilon_{i,t+1}\varepsilon_{m,t+1}^2\right)}{\sqrt{E\left(\varepsilon_{i,t+1}^2\right)}E\left(\varepsilon_{m,t+1}^2\right)}$$

Figure C3: VBA code to calculate conditional coskewness

```
Function CoSkew(ByRef Share_Ret As Range, ByRef Market_Ret As Range)

    T = Share_Ret.Count
    Dim alpha, beta As Double
    ReDim resid(1, 1 To T) As Double
    ReDim resid_m(1 To T, 1) As Double
    ReDim resid2(1, 1 To T) As Double
    ReDim resid_m2(1 To T, 1) As Double
    Dim i As Long
    Dim Market_Mean, Var_e, Var_m As Double

    Market_Mean = WorksheetFunction.Average(Market_Ret)
    x = WorksheetFunction.LinEst(Share_Ret, Market_Ret, True, True)
    beta = x(1, 1)
    alpha = x(1, 2)
```

```
    For i = 1 To T
        resid(1, i) = Share_Ret(i) - alpha - beta * Market_Ret(i)
        resid_m(i, 1) = Market_Ret(i) - Market_Mean
        resid2(1, i) = (Share_Ret(i) - alpha - beta * Market_Ret(i)) ^ 2
        resid_m2(i, 1) = (Market_Ret(i) - Market_Mean) ^ 2
        Var_e = Var_e + resid2(1, i)
        Var_m = Var_m + resid_m2(i, 1)
    Next

    Var_e = Var_e / T
    Var_m = Var_m / T
    temp = M_Mult(resid, resid_m2)
    temp2 = temp(1, 1) / (T * ((Var_e) ^ (1 / 2) * Var_m))
    CoSkew = temp2

End Function
```

Function *CoKurt* calculates conditional cokurtosis, β_i^k, as per Equation 2.27:

$$\beta_i^k = \frac{E\left(\varepsilon_{i,t+1}^2 \varepsilon_{m,t+1}^2\right)}{E\left(\varepsilon_{i,t+1}^2\right) E\left(\varepsilon_{m,t+1}^2\right)}$$

Figure C4: VBA code to calculate unconditional cokurtosis

```
Function CoKurt(ByRef Share_Ret As Range, ByRef Market_Ret As Range)

    T = Share_Ret.Count
    Dim alpha, beta As Double
    ReDim resid(1 To T, 1) As Double
    ReDim resid_m(1 To T, 1) As Double
    ReDim resid2(1, 1 To T) As Double
    ReDim resid_m2(1 To T, 1) As Double
    Dim i As Long
    Dim Market_Mean, Var_e, Var_m As Double



    Market_Mean = WorksheetFunction.Average(Market_Ret)
    x = WorksheetFunction.LinEst(Share_Ret, Market_Ret, True, True)
    beta = x(1, 1)
    alpha = x(1, 2)

    For i = 1 To T
        resid(i, 1) = Share_Ret(i) - alpha - beta * Market_Ret(i)
        resid_m(i, 1) = Market_Ret(i) - Market_Mean
        resid2(1, i) = (Share_Ret(i) - alpha - beta * Market_Ret(i)) ^ 2
        resid_m2(i, 1) = (Market_Ret(i) - Market_Mean) ^ 2
        Var_e = Var_e + resid2(1, i)
        Var_m = Var_m + resid_m2(i, 1)
    Next
    Var_e = Var_e / T
    Var_m = Var_m / T
    temp = M_Mult(resid2, resid_m2)
    temp2 = temp(1, 1) / (T * (Var_e * Var_m))

    CoKurt = temp2

End Function
```

Appendix C: Python Code to Establish Class of Regression

This appendix details the Python 2.7 programming code utilised to calculate ordinary least squares, generalised least squares, and the generalised method of moments estimates and test statistics. The programming converts the estimation formulas outlined in Cochrane (2005) and their multidimensional dependent variable extensions, documented by Siriwardane (2013), into callable functions. The *Regress_dissertation* class is written as a package that can be imported and used in any python program. *Regress_dissertation* objects are initiated with inputs $\boldsymbol{R^e}$, a $T \times N$ vector of portfolio excess returns, and $\boldsymbol{F}$, a $T \times K$ vector of factor returns. *Regress_dissertation* outputs are commented in figure D1.

```
1.  import numpy as np
2.  from numpy import ceil,argsort,zeros, eye,reshape, kron,dot,var,exp,mean,squeeze
3.  from numpy.linalg import lstsq, inv, pinv, solve, eigh
4.  from numpy.linalg import matrix_rank as rank
5.  from pandas import read_csv
6.  from scipy.stats import f, chi2,t
7.  import statsmodels.api as sm
8.  import ENSm as ENS
9.
10. class regress(object):
11.     def __init__(self,y,X):
12.         """
13.         N = number of dependent variables
14.         T = number of data points
15.         k = number of factors
16.         yp = portfolios (not excess returns)
17.         y = dependent variable and has shape (T,N)
18.         X = independent variable and has shape (T,k)
19.         """
20.         # Set shape vars
21.
22.         self.T = y.shape[0]
23.         self.N = y.shape[1]
24.         self.k = X.shape[1]
25.         self.nmom = self.N + self.N*self.k + self.N
26.
27.         # Set endog and exog vars
28.
29.         self.y = y
30.         self.X = X
31.         self.fac = X.T #(the transpose is so we can work with column vectors)
32.
33.         # Calculate vcv of factors to correct for autocorrelation and heteroscedacity
34.
35.         self.fm = reshape(mean(self.fac,axis=1),(self.k,1))
36.         self.vcvfac = dot(self.fac-self.fm,(self.fac - self.fm).T)/(self.T)
37.
38.         # Calculate mean returns for CS regressions
39.
40.         self.ym = y.mean(axis=0)
41.
42.         #Setup identity matrices so we dont have to initialise them every time
43.
44.         self.Ik = eye(self.k)
45.         self.IN = eye(self.N)
46.         self.INT = eye(self.N*self.T)
47.
```

```
48.         # TS regression with an intercept
49.
50.         self.thetaOLS, self.Sigmai, self.uti = self.TS_OLS(const=True)
51.         self.TSalpha = self.thetaOLS[0:self.N]
52.         self.TSbeta = reshape(self.thetaOLS[self.N::],(self.N,self.k)).T
53.
54.         # Fama-Macbeth regression
55.
56.         alpha,lamda,alphat,lambdat,vcvalpha,vcvlamb,R2,R2adj,ts_fmb,ts_pvalue =
57.         self.fmb(self.y,self.TSbeta)
58.         self.tsLamFMB = ts_fmb
59.         self.pvLamFMB = ts_pvalue
60.         self.fmb_pval = self.fmb_pval(alpha,vcvalpha)
61.
62.         # TS regressions without an intercept
63.
64.         self.betaOLS, self.Sigmani, self.utni = self.TS_OLS()
65.         self.Sigma = self.Sigmai
66.         #m = int(ceil(1.2 * float(self.T)**(1.0/3)))
67.         #self.Sigma = self.set_S(m,self.utni.T)  #In case we want a weighting matrix
68.         if (rank(self.Sigma,tol=1e-9) < self.N):
69.             print "Warning Sigma has deficient rank of ", rank(self.Sigma,tol=1e-9)
70.         Bi = self.TSbeta.T
71.         Bni = reshape(self.betaOLS,(self.N,self.k))
72.
73.         # Compute GRS test on TS regression
74.
75.         self.GRStest = self.GRS_test(self.TSalpha,self.Sigma)
76.
77.         # GMM regression with parameters estimated using the standard regression but errors
            corrected for using a HAC matrix
78.
79.         self.do_TS_GMM(self.uti.T,self.utni.T,self.thetaOLS,self.fac)
80.         self.do_CS_GMM(self.uti.T,self.thetaOLS,self.fac,alphat.T,alpha,lamda)
81.
82.         # OLS CS Regressions
83.
84.         self.LambdasOLS, self.alphasOLS, self.covLamOLS, self.covAlpOLS, = self.CS_OLS(Bni)

85.         self.tsLamOLS, self.pvLamOLS = self.get_tstat(lamda, self.covLamOLS)
86.         self.CS_OLS_JS = self.CS_OLS2() [8]
87.         self.CS_OLS_pval = self.CS_OLS2() [9]
88.
89.        # OLS correction for AC and CH
90.
91.         self.covLamOLSC,self.covAlpOLSC = self.correct_ACCH(self.LambdasOLS,self.covLamOLS,
92.         self.covAlpOLS)
93.
94.         # GLS CS Regressions
95.
96.         self.LambdasGLS, self.alphasGLS, self.covLamGLS, self.covAlpGLS = self.CS_GLS(Bni)
97.         self.tsLamGLS, self.pvLamGLS = self.get_tstat(self.LambdasGLS, self.covLamGLS)
98.         self.CS_GLS_JS = self.get_Jstat(self.alphasGLS,self.Sigma,corr = self.T*
99.         (1 + dot(self.LambdasGLS.T,solve(self.vcvfac,self.LambdasGLS)))) [1]
100.
101.         # GLS correction for AC and CH
102.
103.         self.covLamGLSC, self.covAlpGLSC = self.correct_ACCH(self.LambdasGLS,
104.         self.covLamGLS,self.covAlpGLS)
105.
106.     def TS_OLS(self,const=False):
107.         """
108.         This function performs ordinary OLS on the time series. If const = True then an
109.         intercept is included in regressions.
110.         """
111.         if (const==False):
```

```
            # Get data in SUR shape
            XSUR = zeros([self.N*self.T,self.N*self.k])
            for i in range(self.T):
                XSUR[i*self.N:(i+1)*self.N,:] = kron(eye(self.N),reshape(self.X[i,:],
                (1,self.k)))
        elif (const==True):
            # Get data in SUR shape
            XSUR = zeros([self.N*self.T,self.N*(self.k + 1)])
            for i in range(self.T):
                XSUR[i*self.N:(i+1)*self.N,0:self.N] = eye(self.N)
                XSUR[i*self.N:(i+1)*self.N,self.N::] = kron(eye(self.N),
                reshape(self.X[i,:],(1,self.k)))

        # Calculate OLS parameters and residuals

        ylong = reshape(self.y,(self.N*self.T,1))
        theta = lstsq(XSUR,ylong)[0]
        et = ylong - dot(XSUR,theta)
        sigma = var(et)
        Sigma = sigma*self.IN

        # Reshape residuals to get old shape

        ut = reshape(et,(self.T,self.N))
        ut2 = ut.T
        utm = reshape(mean(ut2,axis=1),(self.N,1))
        Sigma = dot(ut2-utm,(ut2 - utm).T)/(self.T)
        return theta, Sigma, ut

    def CS_OLS(self,beta):
        """
        This function runs a cross-sectional OLS regression
        """
        BB = dot(beta.T,beta)
        if (rank(BB,tol=1e-9) < self.k):
            print "Warning, B.T B has deficient rank of '", rank(BB)
            Binv = pinv(dot(beta.T,beta))
        else:
            Binv = inv(dot(beta.T,beta))
        Lamb = dot(Binv,dot(beta.T,self.ym))
        alphas = self.ym - dot(beta,Lamb)

        # Calculate covariance matrices

        covLamb = dot(Binv,dot(dot(dot(beta.T,self.Sigma),beta),Binv))/self.T
        t1 = eye(self.N) - dot(beta,dot(Binv,beta.T))
        covAlp = dot(dot(t1,self.Sigma),t1)/self.T

        return Lamb, alphas, covLamb, covAlp

    def CS_OLS2(self):
        """
        For comparison we do cross-sectional OLS with statsmodels package
        """
        beta = self.TSbeta.T
        vcv = dot(self.uti.T,self.uti)/self.T

        # Run CS regression

        model = sm.OLS(self.ym,beta).fit()
        alpha = model.resid
        Lamb = model.params
        vcvLamb = model.cov_params()
        tstat, pvalues = self.get_tstat(Lamb, vcvLamb)
        R2 = model.rsquared
        R2adj = model.rsquared_adj
```

```
178.
179.         # Wald test
180.
181.         tmp = (self.IN - dot(beta,dot(inv(dot(beta.T,beta)),beta.T)))
182.         vcvalpha = dot(tmp,dot(vcv,tmp))/self.T
183.         Jstat = (self.T-self.N-1)*dot(alpha,dot(pinv(vcvalpha),alpha.T))/self.T
184.         Jpval = 1 - chi2(self.N-self.k).cdf(Jstat)
185.         self.OLS_pval = Jpval
186.
187.         return alpha,Lamb,vcvalpha,vcvLamb,tstat,pvalues,R2,R2adj,Jstat,Jpval
188.
189.     def CS_GLS(self,beta):
190.         """
191.         This function runs a cross-sectional GLS regression
192.         """
193.         if (rank(self.Sigma,tol=1e-9) < self.N):
194.             BSinv = dot(beta.T,pinv(self.Sigma))
195.         else:
196.             BSinv = dot(beta.T,inv(self.Sigma))
197.         tmp = dot(BSinv,beta)
198.         if (rank(tmp,tol=1e-9) < self.k):
199.             print "Warning tmp has deficient rank of ", rank(tmp)
200.             Binv = pinv(tmp)
201.         else:
202.             Binv = inv(tmp)
203.         Lambdas = dot(Binv,dot(BSinv,self.ym))
204.         alphas = self.ym - dot(beta,Lambdas)
205.
206.         # Calculate covariance matrices
207.
208.         covLam = Binv/self.T
209.         covAlp = (self.Sigma - dot(beta,dot(Binv,beta.T)))/self.T
210.         return Lambdas, alphas, covLam, covAlp
211.
212.     def fmb(self,y,beta):
213.         """
214.         This function computes fmb regressions with constant betas
215.         """
216.         # Create array to store results
217.
218.         lambdat = zeros([self.T,self.k])
219.         alphat = zeros([self.T,self.N])
220.         avgPort = mean(self.y, axis=0)
221.
222.         # Run CS regressions
223.
224.         for i in range(self.T):
225.             lambdat[i] = lstsq(beta.T,y[i,:])[0]
226.             alphat[i] = y[i] - dot(lambdat[i].T,beta)
227.
228.         # Calculate alpha & lambda and their covariance matrices
229.
230.         alpha = mean(alphat, axis=0)
231.         lamb = mean(lambdat, axis=0)
232.         vcvalpha = dot((alphat - alpha).T,(alphat - alpha))/self.T**2
233.         vcvlamb = dot((lambdat - lamb).T,(lambdat - lamb))/self.T**2
234.
235.         # Calculate R2, t-statistics and p-values
236.
237.         R2 = 1 - np.var(alpha) / np.var(avgPort)
238.         R2adj = R2 - (1-R2) * self.k / (self.N - self.k -1 )
239.         tstat, pvalue = self.get_tstat(lamb,vcvlamb)
240.
241.         return alpha,lamb,alphat,lambdat,vcvalpha,vcvlamb,R2,R2adj,tstat,pvalue
242.
243.     def tv_fmb(self,window):
```

```
        """
        This function computes fmb regressions with time-varying betas
        """
        #Create array to store results

        tv_T = self.T - window
        lambdat = zeros([tv_T,self.k])
        alphat = zeros([tv_T,self.N])

        #Run TS & CS regressions

        augFactors = np.hstack((np.ones((len(self.y),1)),self.X))
        avgPort = mean(self.y, axis=0)
        for i in xrange(tv_T):
            X_temp = augFactors[i:i+window,:]
            Y_temp = self.y[i:i+window,:]
            out = lstsq(X_temp,Y_temp)
            beta = out[0][1:]
            tv_y = self.y[i,:]
            CS = lstsq(beta.T,tv_y)
            lambdat[i] = CS [0]
            alphat[i] = tv_y - dot(lambdat[i].T,beta)

        # Calculate alpha & lambda and their covariance matrices

        alpha = mean(alphat, axis=0)
        lamb = mean(lambdat, axis=0)
        vcvalpha = dot((alphat - alpha).T,(alphat - alpha))/tv_T**2
        vcvlamb = dot((lambdat - lamb).T,(lambdat - lamb))/tv_T**2

        # Calculate R2, t-statistics and p-values

        R2 = 1 - np.var(alpha) / np.var(avgPort)
        R2adj = R2 - (1-R2) * self.k / (self.N - self.k -1 )
        tstat, pvalue = self.get_tstat(lamb,vcvlamb)

        return alpha,lamb,alphat,lambdat,vcvalpha,vcvlamb,R2,R2adj,tstat,pvalue

    def fmb_pval(self,alpham,vcvalpha):
        if (rank(vcvalpha,tol=1e-9) < self.N):
            self.FMB_JS = dot(alpham.T,dot(pinv(vcvalpha),alpham))
            return chi2.sf(dot(alpham.T,dot(pinv(vcvalpha),alpham)),self.N- self.k)
        else:
            self.FMB_JS = dot(alpham.T,solve(vcvalpha,alpham))
            return chi2.sf(dot(alpham.T,solve(vcvalpha,alpham)),self.N - self.k)

    """
    This function computes GMM functions and finally GMM regressions
    """

    def set_b(self,c,beta,lam = 0.0,method='TS'):
        #create parameter vector
        if method == 'TS':
            b = zeros([self.N*(self.k + 1),1])
        else:
            b = zeros([self.N*(self.k + 1) + self.k,1])
        b[0:self.N,:] = c
        b[self.N:self.N + self.N*self.k,:] = reshape(beta,(self.N*self.k,1))
        if method != 'TS':
            b[self.N + self.N*self.k::,:] = reshape(lam,(self.k,1))
        return b

    def set_g(self,epsilon,f,alpha = 0.0, method='TS'):
        if method == 'TS':
            g = zeros([self.N + self.N*self.k,self.T])
        else:
```

```
            g = zeros([self.N + self.N*self.k + self.N,self.T])
        for i in range(self.T):
            g[0:self.N,i] = epsilon[:,i]
            g[self.N:(self.N + self.N*self.k),i] = kron(epsilon[:,i],self.fac[:,i])
            if method != 'TS':
                g[(self.N + self.N*self.k)::,i] = alpha[:,i]
        return g

    def set_d(self,lambdas=0.0,B = 0.0,method='TS'):
        if method == 'TS':
            d = zeros([self.N + self.N*self.k,self.N + self.N*self.k])
        else:
            d = zeros([self.N*(self.k +1) + self.N,self.N*(self.k +1) + self.k])
        d[0:self.N,0:self.N] = self.IN
        d[0:self.N,self.N:self.N*(self.k+1)] = kron(self.IN,self.fm.T)
        for i in range(self.N):
            n = self.N + i*self.k
            d[n:(n + self.k),0:self.N] = kron(self.IN[i,:],self.fm)
            d[n:(n + self.k),self.N:self.N*(self.k+1)] = kron(self.IN[i,:],self.vcvfac)
        if method != 'TS':
            d[self.N*(self.k+1)::,self.N:self.N*(self.k+1)] = kron(self.IN,lambdas.T)
            d[self.N*(self.k+1)::,self.N*(self.k+1)::] = B
        return -d


    def set_S(self,m,gt,kernel='HAC'):
        nmom = gt.shape[0]
        S = zeros([nmom,nmom])
        Gamma = zeros([m,nmom,nmom])
        #w = zeros([m,m])
        wd = zeros([m])
        for i in range(m):
            #Get weight matrix
            if kernel == 'HAC':
                wd[i] = 1.0 - i/(m+1.0)
            elif (kernel == "ExpSq"):
                wd[i] = exp(-(i + 0.0)**2/(m**2/4))
            #Get Gammas
            Gamma[i] += dot(gt[:,i::],gt[:,:self.T-i].T)/self.T
            #Get S
            S += wd[i]*(Gamma[i] + Gamma[i].T)
        return S

    def get_varb(self,d,S,a=0.0,method='TS'):
        if method=='TS':
            dinv = inv(d)
            return dot(dinv,dot(S,dinv.T))/self.T
        else:
            adinv = inv(dot(a,d))
            varb = dot(adinv,dot(dot(a,S),dot(a.T,adinv.T)))/self.T
            nmom = S.shape[0]
            tmp = eye(nmom) - dot(d,dot(adinv,a))
            vargt = dot(tmp,dot(S,tmp.T))/self.T
            return varb, vargt

    def do_TS_GMM(self,utu,utr,b,f):

        # Unrestricted case where intercepts are included

        alpha = b[0:self.N]
        gtu = self.set_g(utu,f)
        nmom = gtu.shape[0]
        d = self.set_d()
        m = int(ceil(1.2 * float(self.T)**(1.0/3))) #int(floor(self.T**(1.0/4.0)))
        Su = self.set_S(m,gtu)
        SIGMAb = self.get_varb(d,Su)
```

```
        Sigmaalp = SIGMAb[0:self.N,0:self.N]
        if rank(Sigmaalp,tol=1e-9)<self.N:
            valu = dot(alpha.T,dot(pinv(Sigmaalp),alpha))
        else:
            valu = dot(alpha.T,solve(Sigmaalp,alpha))
        self.TS_GMM_pval_u = squeeze(chi2.sf(valu,self.N - self.k))

        # Restricted case with no intercept included

        gtr = self.set_g(utr,f)
        Sr = self.set_S(m,gtr)
        gTr = reshape(mean(gtr,axis=1),(nmom,1))
        if rank(Sr,tol=1e-9) < nmom:
            valr = self.T*dot(gTr.T,dot(pinv(Sr),gTr))
        else:
            valr = self.T*dot(gTr.T,solve(Sr,gTr))
        self.TS_GMM_pval_r = squeeze(chi2.sf(valr,self.N-self.k))

        # GJ test

        gTu = reshape(mean(gtu,axis=1),(nmom,1))
        val = self.T*(dot(gTr.T,solve(Su,gTr)) - dot(gTu.T,solve(Su,gTu)))
        self.TS_GMM_pval_3 = squeeze(chi2.sf(val,self.N-self.k))

        return

    def do_CS_GMM(self,ut,theta,f,alphat,alpha,lamda):

        # TS intercepts & betas

        c = theta[0:self.N]

        beta = theta[self.N::]
        B = reshape(beta,(self.N,self.k))

        # Setup parameter vector

        b = self.set_b(c,beta,lamda,method='CS')
        npar = b.size

        #Setup moments

        gt = self.set_g(ut,f,alphat,method='CS')
        nmom = gt.shape[0]

        # Setup d matrix

        d = self.set_d(lambdas=lamda,B=B,method='CS')

        #Calculate lag

        m = int(ceil(1.2 * float(self.T)**(1.0/3))) #int(floor(self.T**(1.0/4.0)))

        # Calcuate S matrix
        #S = self.set_S(m,gt)  #,kernel="ExpSq"
        #S = S + 1e-8*eye(nmom) #Add jitter since S is near singular

        S = ENS.GMM_cross_sectional(self.y.T,self.X.T) [5]
        self.S = S

        #Setup a matrix

        a = zeros([self.N*(self.k+1) + self.k,self.N*(self.k+1) + self.N])
        a[0:self.N*(self.k+1),0:self.N*(self.k+1)] = eye(self.N*(self.k+1))
        a[self.N*(self.k+1)::,self.N*(self.k+1)::] = B.T
        self.a = a
```

```
442.         self.d = d
443.
444.         #Calculate variance-covariance matrices
445.
446.         vcvb, vcvgt= self.get_varb(d,S,a=a,method='CS')
447.         n = nmom-self.N
448.         Sigmaalp = vcvgt[n::,n::]
449.         self.Sigmaalp = Sigmaalp
450.         vcvLamb = vcvb[-self.k:,-self.k:]
451.         self.covLamGMM = vcvLamb
452.         tstats, pvalues = self.get_tstat(lamda,vcvLamb)
453.         self.tsLamGMM = tstats
454.         self.pvLamGMM = pvalues
455.         # Wald test
456.
457.         if rank(Sigmaalp,tol=1e-9)<self.N:
458.             val = dot(alpha.T,dot(pinv(Sigmaalp),alpha))
459.         else:
460.             val = dot(alpha.T,solve(Sigmaalp,alpha))
461.         self.CS_GMM_pval1 = squeeze(chi2.sf(val,self.N-self.k))
462.
463.         # Test of overidentifying restrictions
464.
465.         gT = reshape(mean(gt,axis=1),(nmom,1))
466.         self.gT = gT
467.         if rank(vcvgt,tol=1e-9) < nmom:
468.             val2 = self.T*dot(gT.T,dot(pinv(S),gT))
469.         else:
470.             val2 = self.T*dot(gT.T,solve(S,gT))
471.
472.         self.CS_GMM_JS = squeeze(val2)
473.         self.CS_GMM_pval2 = squeeze(chi2.sf(val2,nmom-npar))
474.
475.         return
476.
477.     """
478.     This section computes global functions
479.     """
480.
481.     def get_tstat(self, parameter, vcvparameter):
482.         """
483.         This function computes t statistic of the regression coefficients
484.         """
485.         t_parameter = (parameter.T/np.sqrt(np.diag(vcvparameter))).T
486.         pvalues = (1-t.cdf(t_parameter, self.N-self.k))*2
487.
488.         return t_parameter, pvalues
489.
490.     def get_pseudo_R2(self,vcvport,vcvres):
491.         """
492.         Here we get the pseudo R squared statistic
493.         """
494.         # Calculate eigenvals and eigenvecs of portfolio covariance matrix
495.
496.         lam_port, p_port = self.do_spec_dec(vcvport)
497.
498.         # Calculate eigenvals and eigenvecs of residual covariance matrix
499.
500.         lam_res, p_res = self.do_spec_dec(vcvres)
501.
502.         return 1 - sum(lam_res)/sum(lam_port)
503.
504.     def do_spec_dec(self,A):
505.         """
506.         Here we do the spectral decomposition of the square symmetric matrix A
507.         A is decomposed into A = P V P' where P contains the eigenvectors and V
```

```
        is a diagonal matrix containing the eigenvalues of A. The returned eigenvalues
        are ranked in decending order and the ith column of Ps correspondes to the ith
        eigenvalue.
        """
        # Eigen-decomposition of A (this is only for symmetric matrices, for non-
        symmetric matrices use eig instead of eigh)

        V, P = eigh(A)

        # Calculate indices that sort the eigenvals

        I = argsort(V)

        # Sort eigenvecs

        Ps = P[:,I]

        # Sort eigenvals

        Vs = V[I]

        return Vs, Ps

    def correct_ACCH(self,Lambdas,covLam,covAlp):
        """
        Here we correct for the AC and CH.
        """
        # Calculate the multiplicative correction factor

        if (rank(self.vcvfac,tol=1e-9) < self.k):
            print "Warning vcvfac has deficient rank of ", rank(self.vcvfac)
            mcor = (1 + dot(Lambdas.T,dot(pinv(self.vcvfac),Lambdas)))
        else:
            mcor = (1 + dot(Lambdas.T,solve(self.vcvfac,Lambdas)))

        # Correct the alpha and Lambda

        covLamC = covLam*mcor + self.vcvfac/self.T
        covAlpC = covAlp*mcor

        return covLamC, covAlpC

    def GRS_test(self,alpha,Sigma):
        """
        This test should be used for time series regressions
        """
        if (rank(Sigma,tol=1e-9) < self.N):
            print "Warning Sigma has deficient rank of ", rank(Sigma)
            val = (self.T - self.N - self.k)*dot(alpha.T,dot(pinv(Sigma),alpha))/
            (self.N*(1 + dot(self.fm.T,solve(self.vcvfac,self.fm))))
        else:
            val = (self.T - self.N - self.k)*dot(alpha.T,solve(Sigma,alpha))/
            (self.N*(1 + dot(self.fm.T,solve(self.vcvfac,self.fm))))
        return squeeze(f.sf(val,self.N,self.T-self.N-self.k))


    def get_Jstat(self,alpha,sigma,corr = 1.0):
        """
        J test. corr is the prefactor
        """
        if (rank(sigma,tol=1e-9) < self.N):
            print "Warning Sigma has deficient rank of ", rank(sigma)
            val = corr*dot(alpha.T,dot(pinv(sigma),alpha))
        else:
            val = corr*dot(alpha.T,solve(sigma,alpha))
        return chi2.sf(val,self.N-self.k), val
```

```
574.
575.    def CS_OLS_pval_J(self):
576.        """
577.        J test. corr is the prefactor
578.        """
579.        return self.get_Jstat(self.alphasOLS,self.Sigma,corr = self.T) [0]
580.
581.    def CCS_OLS_pval_J(self):
582.        """
583.        J test. corr is the prefactor
584.        """
585.        return self.get_Jstat(self.alphasOLS,self.Sigma,corr = self.T*
586.        (1 + dot(self.LambdasGLS.T,solve(self.vcvfac,self.LambdasGLS)))) [0]
587.
588.    def CS_GLS_pval_J(self):
589.        """
590.        J test. corr is the prefactor
591.        """
592.        return self.get_Jstat(self.alphasGLS,self.Sigma,corr = self.T) [0]
593.
594.    def CCS_GLS_pval_J(self):
595.        """
596.        J test. corr is the prefactor
597.        """
598.        return self.get_Jstat(self.alphasGLS,self.Sigma,corr = self.T*
599.        (1 + dot(self.LambdasGLS.T,solve(self.vcvfac,self.LambdasGLS)))) [0]
```

Appendix D: Python Code for Monte Carlo Experiment

This appendix details the Python 2.7 programming code utilised to analyse size and power properties of the test statistics used within the paper. The *MC_dissertation* program initiates Monte Carlo simulations specifically designed to assess the impact of skewness and kurtosis on hypothesis tests called from *Regress_dissertation*. For each simulation we construct a series of null returns, $\widetilde{\boldsymbol{R}}^{\boldsymbol{n}}$, and a series of alternative returns, $\widetilde{\boldsymbol{R}}^{\boldsymbol{a}}$ as described in the paper. Error terms are drawn from: normal distributions; $\chi^2_{(4)}$ distributions, and central $t_{(3)}$ & central $t_{(5)}$ distributions to investigate the comparative effect of skewness and kurtosis on the selected test statistics.

Results are graphed using the P-value plots and Size-Power curves developed by Davidson and Mackinnon (1997). The graphs are based on the empirical p-value distribution of the given test statistics. The p-values of test statistics J_{OLS}, J_{GLS}, J_{FM} and J_{GMM} are the probability that portfolio pricing errors are jointly zero, i.e. H_o: $\alpha = 0$. At any point x_i falls between zero and one and $\hat{F}(x_i)$ is defined by:

$$\hat{F}(x_i) = \frac{1}{S}\sum_{j=1}^{S} I\,(p_j < x_i)$$

Where S is the number of simulations and $I\,(p_j < x_i)$ is an indicator function that takes a value of 1 if the argument is true or alternatively 0. Although it is possible to evaluate S at every data point, DM suggest we set $x_i = 0.001, 0.002, \ldots, 0.999 \quad (m = 1000)$ to ensure that plotted lines are not jagged.

```
1.  import numpy as np
2.  from numpy import dot,mean,savetxt
3.  from numpy.random import standard_t, standard_normal, chisquare
4.  from numpy.linalg import lstsq
5.  from pandas import read_csv
6.  from regress_dissertation import regress
7.  import matplotlib as mpl
8.  import matplotlib.pyplot as plt
9.
10.
11. # Import data portfolio and factor data
12.
13. data = read_csv('C:\users\Johan\documents\FinalData2.csv')
14. mark_ex= data['Mark_Ex'].values
15. factors = data[['Mark_Ex','SMB','HML']].values
16. riskfree = data['RF'].values
17. portfolios = data[['S1','S2','S3','S4','S5','S6','S7','S8','S9','S10']].values
18.
19. # Setup excess return matrix
20.
21. T,N = portfolios.shape
22. T,K = factors.shape
23. portfolios = portfolios[:,np.arange(0,N)]
24. excessRet = portfolios - np.reshape(riskfree,(T,1))
25.
26. # Setup parameter and residual vectors for simulations
27.
28. X = np.hstack((np.ones((T, 1)), factors))
29. thetaOLS = lstsq(X, excessRet)
```

```
30.  betaOLS = thetaOLS[0][1:]
31.  alpha_a = 0.015
32.  e = excessRet-dot(factors,betaOLS)
33.  residStd = e.std()
34.
35.  # Simulate data with specific error term and perform test statistics
36.
37.  S = 1000
38.  norm_rn = np.zeros([T, N])
39.  norm_ra = np.zeros([T, N])
40.  t3_rn = np.zeros([T, N])
41.  t3_ra = np.zeros([T, N])
42.  t5_rn = np.zeros([T, N])
43.  t5_ra = np.zeros([T, N])
44.  pvalues_n = np.mat(np.zeros([16,S]))
45.  pvalues_a = np.mat(np.zeros([16,S]))
46.  for s in range(S):
47.      # Progress report
48.      print s
49.      # Simulate data
50.      tmp = residStd*standard_normal(size=(T,N))
51.      norm_rn = dot(factors,betaOLS) + tmp
52.      norm_ra = alpha_a + dot(factors,betaOLS) + tmp
53.      tmp = residStd*standard_t(3,size=(T,N))
54.      t3_rn = dot(factors,betaOLS)+ tmp
55.      t3_ra = alpha_a + dot(factors,betaOLS) + tmp
56.      tmp = residStd*standard_t(5,size=(T,N))
57.      t5_rn = dot(factors,betaOLS) + tmp
58.      t5_ra = alpha_a + dot(factors,betaOLS)+ tmp
59.      chi2t = residStd*chisquare(4,size=(T,N))
60.      chi2t = chi2t - mean(chi2t)
61.      chi2_rn = dot(factors,betaOLS) + chi2t
62.      chi2_ra = alpha_a + dot(factors,betaOLS)+ chi2t
63.
64.      # Init regress objects
65.      # For Null
66.      nntest = regress(norm_rn,factors)
67.      nt3test = regress(t3_rn,factors)
68.      nt5test = regress(t5_rn,factors)
69.      nc2test = regress(chi2_rn,factors)
70.
71.      # For alternative
72.      antest = regress(norm_ra,factors)
73.      at3test = regress(t3_ra,factors)
74.      at5test = regress(t5_ra,factors)
75.      ac2test = regress(chi2_ra,factors)
76.
77.
78.      # Null hypothesis
79.      pvalues_n[0,s] = nntest.CS_OLS_pval
80.      pvalues_n[1,s] = nt3test.CS_OLS_pval
81.      pvalues_n[2,s] = nt5test.CS_OLS_pval
82.      pvalues_n[3,s] = nc2test.CS_OLS_pval
83.      pvalues_n[4,s] = nntest.CCS_GLS_pval_J()
84.      pvalues_n[5,s] = nt3test.CCS_GLS_pval_J()
85.      pvalues_n[6, s] = nt5test.CCS_GLS_pval_J()
86.      pvalues_n[7,s] = nc2test.CCS_GLS_pval_J()
87.      pvalues_n[8,s] = nntest.fmb_pval
88.      pvalues_n[9,s] = nt3test.fmb_pval
89.      pvalues_n[10,s] = nt5test.fmb_pval
90.      pvalues_n[11,s] = nc2test.fmb_pval
91.      pvalues_n[12,s] = nntest.CS_GMM_pval2
92.      pvalues_n[13,s] = nt3test.CS_GMM_pval2
93.      pvalues_n[14,s] = nt5test.CS_GMM_pval2
94.      pvalues_n[15,s] = nc2test.CS_GMM_pval2
95.
```

```
96.
97.      #alternate hypothesis
98.       pvalues_a[0,s] = antest.CS_OLS_pval
99.       pvalues_a[1,s] = at3test.CS_OLS_pval
100.      pvalues_a[2,s] = at5test.CS_OLS_pval
101.      pvalues_a[3,s] = ac2test.CS_OLS_pval
102.      pvalues_a[4,s] = antest.CCS_GLS_pval_J()
103.      pvalues_a[5,s] = at3test.CCS_GLS_pval_J()
104.      pvalues_a[6, s] = at5test.CCS_GLS_pval_J()
105.      pvalues_a[7,s] = ac2test.CCS_GLS_pval_J()
106.      pvalues_a[8,s] = antest.fmb_pval
107.      pvalues_a[9,s] = at3test.fmb_pval
108.      pvalues_a[10,s] = at5test.fmb_pval
109.      pvalues_a[11,s] = ac2test.fmb_pval
110.      pvalues_a[12,s] = antest.CS_GMM_pval2
111.      pvalues_a[13,s] = at3test.CS_GMM_pval2
112.      pvalues_a[14,s] = at5test.CS_GMM_pval2
113.      pvalues_a[15,s] = ac2test.CS_GMM_pval2
114.
115.
116.  savetxt('C:\users\Johan\documents\pvaluesn.txt', pvalues_n)
117.  savetxt('C:\users\Johan\documents\pvaluesa.txt', pvalues_a)
118.
119.  #setup vector X and F(x) following Davidson & MacKinnon
120.
121.  V, W = np.shape(pvalues_n)
122.  W = int(S/10)
123.  X = np.linspace(0.001,0.999,W)
124.  fx_n = np.zeros([V,W])
125.  fx_a = np.zeros([V,W])
126.
127.  for i in range(V):
128.      for j in range(W):
129.          fx_n[i,j] = np.argwhere(pvalues_n[i,:] <= X[j]).shape[0]
130.          fx_a[i,j] = np.argwhere(pvalues_a[i,:] <= X[j]).shape[0]
131.
132.  fx_n = (fx_n/S).T
133.  fx_a = (fx_a/S).T
134.  T = str(T)
135.  N = str(N)
136.  K = str(K)
137.
138.  # Plot p-value curves
139.
140.  fig, ax = plt.subplots(nrows = 2, ncols = 2)
141.  fig.suptitle('P-Value Plots (T = ' + T + ', N = ' +
142.   N + ', K = ' + K + ')',fontsize=20)
143.  mpl.rcParams.update({'font.size': 10, 'font.family': 'serif'})
144.  plt.rc('legend',**{'fontsize':8})
145.
146.  # First axis
147.  ax[0,0].set_title('Normal Data')
148.  ax[0,0].plot(X[:], X[:], 'k',label="45deg line")
149.  ax[0,0].plot(X[:],fx_n[:,0],'r',label="J_OLS")
150.  ax[0,0].plot(X[:], fx_n[:,4],'g',label="J_GLS")
151.  ax[0,0].plot(X[:],fx_n[:,8],'b',label="J_FM")
152.  ax[0,0].plot(X[:],fx_n[:,12],'c',label="J_GMM")
153.  ax[0,0].set_xlim([0, 0.15])
154.  ax[0,0].set_ylim([0, 0.15])
155.  ax[0,0].legend(loc=4)
156.
157.  # Second axis
158.  ax[0,1].set_title('Student t(3) Data')
159.  ax[0,1].plot(X[:], X[:], 'k',label="45deg line")
160.  ax[0,1].plot(X[:],fx_n[:,1],'r',label="J_OLS")
161.  ax[0,1].plot(X[:], fx_n[:,5],'g',label="J_GLS")
```

```
ax[0,1].plot(X[:],fx_n[:,9],'b',label="J_FM")
ax[0,1].plot(X[:],fx_n[:,13],'c',label="J_GMM")
ax[0,1].set_ylim([0, 0.15])
ax[0,1].legend(loc=4)

# Third axis
ax[1,0].set_title('Student t(5) Data')
ax[1,0].plot(X[:], X[:], 'k',label="45deg line")
ax[1,0].plot(X[:],fx_n[:,2],'r',label="J_OLS")
ax[1,0].plot(X[:], fx_n[:,6],'g',label="J_GLS")
ax[1,0].plot(X[:],fx_n[:,10],'b',label="J_FM")
ax[1,0].plot(X[:],fx_n[:,14],'c',label="J_GMM")
ax[1,0].set_xlabel('Nominal Size', fontsize=10)
ax[1,0].set_ylabel('Actual Size', fontsize=10)
ax[1,0].set_xlim([0, 0.15])
ax[1,0].set_ylim([0, 0.15])
ax[1,0].legend(loc=4)

# Fourth axis
ax[1,1].set_title('Chi-Sqaured Data')
ax[1,1].plot(X[:], X[:], 'k',label="45deg line")
ax[1,1].plot(X[:],fx_n[:,3],'r',label="J_OLS")
ax[1,1].plot(X[:], fx_n[:,7],'g',label="J_GLS")
ax[1,1].plot(X[:],fx_n[:,11],'b',label="J_FM")
ax[1,1].plot(X[:],fx_n[:,15],'c',label="J_GMM")
ax[1,1].set_xlabel('Nominal Size', fontsize=10)
ax[1,1].set_xlim([0, 0.15])
ax[1,1].set_ylim([0, 0.15])
ax[1,1].legend(loc=4)

# Plot size-power urves

figsp, axsp = plt.subplots(nrows = 2, ncols = 2)
figsp.suptitle('Size-Power Plots (T = ' + T + ', N = ' +
 N + ', K = ' + K + ')', fontsize=20)
mpl.rcParams.update({'font.size': 10, 'font.family': 'serif'})
plt.rc('legend',**{'fontsize':8})

# First axis
axsp[0,0].set_title('Normal Data')
axsp[0,0].plot(fx_n[:,0],fx_a[:,0],'r',label="J_OLS")
axsp[0,0].plot(fx_n[:,4],fx_a[:,4],'g',label="J_GLS")
axsp[0,0].plot(fx_n[:,8],fx_a[:,8],'b',label="J_FM")
axsp[0,0].plot(fx_n[:,12],fx_a[:,12],'c',label="J_GMM")
axsp[0,0].set_ylabel('Power', fontsize=10)
axsp[0,0].set_xlim([0, 1])
axsp[0,0].set_ylim([0, 1])
axsp[0,0].legend(loc=4)

# Second axis
axsp[0,1].set_title('Student t(3) Data')
axsp[0,1].plot(fx_n[:,1],fx_a[:,1],'r',label="J_OLS")
axsp[0,1].plot(fx_n[:,5],fx_a[:,5],'g',label="J_GLS")
axsp[0,1].plot(fx_n[:,9],fx_a[:,9],'b',label="J_FM")
axsp[0,1].plot(fx_n[:,13],fx_a[:,13], 'c',label="J_GMM")
axsp[0,1].set_xlim([0, 1])
axsp[0,1].set_ylim([0, 1])
axsp[0,1].legend(loc=4)

# Third axis
axsp[1,0].set_title('Student t(5) Data')
axsp[1,0].plot(fx_n[:,2],fx_a[:,2],'r',label="J_OLS")
axsp[1,0].plot(fx_n[:,6],fx_a[:,6],'g',label="J_GLS")
axsp[1,0].plot(fx_n[:,10],fx_a[:,10],'b',label="J_FM")
axsp[1,0].plot(fx_n[:,14],fx_a[:,14], 'c',label="J_GMM")
axsp[1,0].set_xlabel('Size', fontsize=10)
```

```
axsp[1,0].set_ylabel('Power', fontsize=10)
axsp[1,0].set_xlim([0, 1])
axsp[1,0].set_ylim([0, 1])
axsp[1,0].legend(loc=4)

# Fourth axis
axsp[1,1].set_title('Chi-Sqaured Data')
axsp[1,1].plot(fx_n[:,3],fx_a[:,3],'r',label="J_OLS")
axsp[1,1].plot(fx_n[:,7],fx_a[:,7],'g',label="J_GLS")
axsp[1,1].plot(fx_n[:,11],fx_a[:,11],'b',label="J_FM")
axsp[1,1].plot(fx_n[:,15],fx_a[:,15], 'c',label="J_GMM")
axsp[1,1].set_xlabel('Size', fontsize=10)
axsp[1,1].set_xlim([0, 1])
axsp[1,1].set_ylim([0, 1])
axsp[1,1].legend(loc=4)
```