

The copyright of this thesis vests in the author. No quotation from it or information derived from it is to be published without full acknowledgement of the source. The thesis is to be used for private study or non-commercial research purposes only.

Published by the University of Cape Town (UCT) in terms of the non-exclusive license granted to UCT by the author.

Dynamic and Robust Estimation of Risk and Return in Modern Portfolio Theory



A thesis presented by

Rodwel Mupambirci

In partial fulfillment of the requirements for the degree of

Master in Science

by coursework and dissertation

in the specialization of

Mathematics of Finance

supervised by

Professor C Troskie

University of Cape Town

August 2008

Abstract

The portfolio selection method developed by Markowitz gives a rational investor a way of evaluating different investment options in a portfolio using the expected return and variance of the returns. Sharpe uses the same optimization approach but estimates the mean and covariance in a regression framework using the index models. Sharpe makes a crucial assumption that the residuals from different assets are uncorrelated and that the beta estimates are constant. When the Sharpe model parameters are estimated using ordinary least squares, the regression assumptions are violated when there is significant autocorrelation and heteroskedasticity in the residuals. Furthermore, the presence of outlying observations in the data leads to unreliable estimates when the ordinary least squares method is used.

We find significant correlation in the residuals from different shares and thus we use the Troskie-Hossain model which relaxes this assumption and ultimately produces an efficient frontier that is almost identical to the Markowitz model. The combination of the GARCH and AR models to remove both autocorrelation and heteroskedasticity is used on the single index model and it causes the efficient frontier to shift significantly to the left. Using dynamic estimation through the Kalman filter, it is noticed that the beta coefficients are not constant and that the resulting efficient frontiers significantly outperform the Sharpe model.

In order to deal with the problem of outlying observations in the data, we propose using the Minimum Covariance Determinant, (MCD) estimator as a robust version of the Markowitz formulation. Robust alternatives to the ordinary least squares estimator are also investigated and they all cause the efficient frontier to shift to the left. Finally, to solve the problem of collinearity in the multiple index framework, we construct orthogonal indices using principal components regression to estimate the efficient frontier.

Acknowledgments

I would like to thank the following people for their help in this thesis.

- My supervisor Professor Troskie for his guidance and support.
- My mother for her encouragement and wisdom.
- The Department of Statistical Sciences for providing stock market data.
- Sabine Verboven and Mia Hubert for permission to use the LIBRA MATLAB toolbox for robust estimation.

Contents

1	Introduction	8
1.1	Subject background	8
1.2	Objectives of the Study	10
1.3	Scope and Limitations	11
1.4	Plan of Development	12
2	Markowitz Portfolio Theory	13
2.1	Assumptions of the Markowitz Model	13
2.2	Markowitz Formulation	14
2.3	The Efficient Frontier	15
2.4	The Optimal Portfolio	16
2.5	Empirical Investigation of the Markowitz Model	18
2.5.1	Objectives and background of the Investigation	18
2.5.2	Data used in the investigation	18
2.5.3	Method of Investigation	20
2.5.4	Results	20
2.6	Analysis of the Results and Conclusion	22
3	Index Models of Asset Returns	23
3.1	The Bivariate Normal Distribution	23
3.2	The Sharpe Single Index Model	24
3.3	Sharpe Single Index model in Portfolio Theory	25
3.4	Empirical Investigation-Sharpe Single Index Model	28
3.4.1	Objectives of the Investigation	28

3.4.2	Data used in the Investigation	28
3.4.3	Method of Investigation	28
3.4.4	Results	29
3.4.5	Conclusion	31
3.5	The Troskie-Hossain Improved Sharpe Model	31
3.6	Empirical Investigation of the Troskie-Hossain Improved Sharpe Single Index Model	34
3.6.1	Objectives of the Investigation	34
3.6.2	Methodology	34
3.6.3	Results	35
3.6.4	Conclusion	39
3.7	The Sharpe Multiple Index Models	40
3.8	The Troskie-Hossain Improved Multiple Index Model	43
3.9	Empirical Investigation of the Multiple Index Models	45
3.9.1	Objectives of the Investigation	45
3.9.2	Data used in the investigation	45
3.9.3	Methodology	45
3.9.4	Results	47
3.9.5	Conclusions	51
3.10	Time Series Errors and Heteroskedasticity-Dynamic Models	52
3.10.1	Autoregressive Models	53
3.10.2	GARCH(p, q) Models	53
3.11	Empirical Investigation of the AR and GARCH Index Models	54

3.11.1	Objectives of the Investigation	54
3.11.2	Data used in the Investigation	55
3.11.3	Method of Investigation	55
3.11.4	Results	55
3.11.5	Conclusions	58
4	State Space Models with the Kalman Filter	60
4.1	State Space Analysis	60
4.2	The Kalman Filtering Algorithm	60
4.3	Rationale behind the Kalman Filter	62
4.4	Estimation of the Parameters	63
4.5	Smoothed State Estimates	64
4.6	Application of The Kalman Filter to the Single Index Models .	65
4.7	The Dynamic Mean Model	68
4.8	Empirical Investigation of The Kalman Filter in Portfolio Theory	69
4.8.1	Objectives of the Investigation	69
4.8.2	Method of Investigation	69
4.8.3	Results	71
4.9	Conclusions	78
5	Robust Portfolio Selection Techniques	80
5.1	Robustness and its Importance in Estimation	80
5.2	Multivariate Robust Estimation Theory	81

5.2.1	The Influence Function	81
5.2.2	The Breakdown Point of an Estimator	84
5.2.3	Affine Equivariance	86
5.2.4	Consistency and Efficiency	86
5.3	The Minimum Covariance Determinant Estimator	88
5.4	Empirical Investigation of the Minimum Covariance Determinant Estimator	90
5.4.1	Objectives of the Investigation	90
5.4.2	Data used in the Investigation	91
5.4.3	Method of Investigation	91
5.4.4	Results	92
5.4.5	Conclusion	98
5.5	Robust Regression Techniques	98
5.5.1	Least Squares Regression	99
5.5.2	Least Trimmed Squares Estimator	100
5.5.3	Empirical Investigation on the Least Trimmed Squares Estimator	101
5.5.4	L_p Norm Estimation	104
5.5.5	Empirical investigation on the L_1 Regression Estimator	104
5.6	Iteratively-reweighted Least Squares Regression	106
5.6.1	M-estimation in Robust Regression	106
5.6.2	Robust Weighting Functions	107
5.6.3	Iteratively Reweighting Algorithm	110
5.7	Empirical Investigation on Iteratively-Reweighted Least Squares	

Estimator	112
5.7.1 Objectives of the Investigation	112
5.7.2 Method of Investigation	112
5.7.3 Results	112
5.8 Conclusions	115
6 Principal Components Analysis	117
6.1 Theory of Principal Components Analysis	117
6.2 Principal Components and Regression	118
6.2.1 Dynamic Principal Components Regression	119
6.2.2 Principal Components Kalman Filtering	121
6.3 Empirical Investigation of Principal Components Analysis . .	122
6.3.1 Objectives of the Investigation	122
6.3.2 Data used in the investigation	122
6.3.3 Method of Investigation	123
6.3.4 Results- Principal Components Regression	124
6.3.5 Results - Dynamic Principal Components Regression .	125
6.3.6 Results- Principal Components Kalman Filtering . . .	128
6.4 Conclusions	130
7 Conclusions	131
7.1 Future Research	133

1 Introduction

1.1 Subject background

In his landmark paper entitled "Portfolio Selection", Harry Markowitz (1952) showed how an investor faced with the problem of selecting an optimal portfolio can use the mean and the covariance structure of the asset returns to construct the efficient frontier. Using the Markowitz mean-variance formulation, the portfolio managers' problem of optimal portfolio selection reduces to selecting a portfolio that lies on the efficient frontier. Although several other portfolio optimization techniques have been proposed since Markowitz proposed his method, the Markowitz formulation remains the basis for many of the newer techniques. Although the optimization process is relatively straightforward, the central problem in portfolio theory is estimating the inputs. Sharpe estimates the inputs through the single and multiple index regression models. Troskie, Hossain and Guo (2005) further refine the index models through the Troskie-Hossain model by improving on the estimation of the covariance structure. Regression models however make strong assumptions about the estimated parameters. For example in the single index model framework, if R_{it} represents the return from the i -th share and I_t is the return on the market or some other market proxy at time t , then Sharpe, (1970), uses the following relationship between the returns:

$$R_{it} = \alpha_i + \beta_i I_t + e_{it} \tag{1.1}$$

with

$$\begin{aligned}
E(e_{it}) &= 0, \quad i = 1, \dots, p \\
E(e_{it}^2) &= \sigma_{ei}^2 \\
E(e_{it}e_{is}) &= 0, \quad t \neq s \\
E(e_{it}I_t) &= 0, \quad t = 1, \dots, N, \\
E(e_{it}e_{jt}) &= 0, \quad i \neq j
\end{aligned} \tag{1.2}$$

From the above formulation, it is clear that Sharpe assumes that the residuals between the stock returns are uncorrelated and there is an assumption that the regression coefficient is constant. For a portfolio manager who uses the Sharpe formulation, these assumptions need to be tested before the model is implemented. From a risk measurement point of view, it is particularly important to investigate the assumptions that the estimates of the betas and the residual variance are constant because these are the critical inputs that contribute to portfolio risk. If the beta coefficients are in fact not constant, then the portfolio manager would like to use the most recent estimates of the betas since these reflect the most current market conditions.

The models used in portfolio theory estimate the parameters from historical data. The problem with financial data is the effect that outliers and influential observations can have on the parameter estimates. The challenge we face is to find estimation methods that are *robust*. Several robust procedures have been proposed by Huber (1981), and Rousseeuw and Leroy (1987) and these methods have shown to be very effective particularly in a regression setting.

In financial data, we are also interested in identifying the factors that explain the variation in the share returns. This is very useful for portfolio management because it can help in understanding the covariance structure of the returns which is a critical input in the portfolio optimization process.

Principal components analysis lends itself to be a very useful tool for this, particularly due to the fact that it requires less restrictive assumptions than other regression type models.

1.2 Objectives of the Study

The Markowitz and Sharpe models make a number of restrictive assumptions particularly regarding the constant parameter estimates. We would like to use models that are dynamic in the sense that they can adapt to changes in the underlying parameters over time. Dynamic models should in theory be expected to give us results that are more useful in practice because by allowing parameter estimates to change with time, we are able to use time as an additional explanatory variable for variation, and since time is predictable, we should be able to reduce the residual risk. More precisely, the objectives of this thesis are to:

- Evaluate the Markowitz formulation as a tool for asset allocation and critically analyze the underlying assumptions that are used in developing this model.
- Review the theory of the Kalman Filter and evaluate its usefulness in dynamic estimation of the mean and covariance structure of asset returns.
- Contrast the optimal asset allocations that result from fitting the Sharpe single and multiple index models, the Kalman Filtered and the Troskie-Hossain, (Hossain, Troskie and Guo, 2006) improvements to Sharpe. Particular attention will be paid to investigating how the efficient frontier shifts in the different cases. This will allow us to contrast how the optimal asset allocation differs in each case.
- Investigate the role that principal components analysis can play in portfolio theory using the Sharpe multiple index model framework through

principal components regression. In addition, the idea of Dynamic Principal Components will be used in order to compare the effect it has on risk reduction in portfolio optimization.

- Introduce models that are robust in the sense that the parameter estimates are not greatly influenced by influential observations. As Huber, (1981, pg1) puts it, "...robustness signifies insensitivity to small deviations from the assumptions." This is precisely what we would like to achieve because we will be using models that depend on certain assumptions, and we would like to see what improvements to the estimation of risk we can get from using robust estimation.

1.3 Scope and Limitations

The results we will get from using various models in this thesis depend on certain statistical assumptions regarding the underlying distributions, independence, autocorrelation and homoskedasticity. Although the data we use may not satisfy some of these assumptions, we will still cautiously use the models assuming these assumptions are not seriously violated. In other words, we expect that these minor violations should not have a serious impact on the financial interpretation of the results. The thesis will try to use a wide variety of statistical models, each of which has a lot of underlying theory and mathematics, some of which is beyond the scope of this thesis. Robust estimation theory is a typical example where this thesis will only be able to state and use the main results without proving much of the underlying theory. The reader is directed to the references for a more rigorous treatment.

1.4 Plan of Development

The thesis starts by reviewing Modern Portfolio Theory using the Markowitz framework in Chapter 2. This is fundamental because our goal throughout this thesis is to construct the efficient frontier that an investor will face when trying to choose the optimal portfolio. In Chapter 3 we introduce the single and multiple index models from Sharpe (1970) and the Troskie-Hossain Improved Sharpe models. We will also discuss the issue of autocorrelation and heteroskedasticity that is often problematic in regression models. In Chapter 4 we introduce the Kalman filter and then we use state space methods to estimate the regression coefficients from the index models. We will then use the estimates derived from the Kalman filter to construct efficient frontiers. Chapter 5 looks at various methods of robust estimation that can be used in robust portfolio optimization. Finally, Chapter 6 looks at the idea of principal components and we investigate the role that state space and autoregressive models can help in producing dynamic principal components.

2 Markowitz Portfolio Theory

2.1 Assumptions of the Markowitz Model

The Markowitz portfolio selection algorithm gives us a powerful tool for constructing the efficient frontier. Given that different assets have different characteristics, the Markowitz Nobel Prize winning formulation gives a principle that can be applied in choosing the best possible combinations of asset weights in a portfolio. The key concept that is used in developing the theory comes from the idea that a rational investor who bases investment decisions on the expected return and variance of asset returns would want a portfolio that provides the highest level of expected return for a given level of risk. The measure of risk that the investor uses is the covariance structure of the asset returns. The assumptions of the Markowitz frameworks are as follows:

- Investors evaluate portfolios by considering only the mean and variance of returns.
- This is done over the same one-period time horizon.
- Investors are rational, in other words they prefer a portfolio with a higher expected return to one with a lower expected return for a given level of risk.
- Investors are risk averse because they prefer a portfolio that has a lower variance to one with a higher variance for a given level of return.
- Individual securities are infinitely divisible such that we can hold them in any proportion of the portfolio.
- Taxes and transaction costs can be ignored.

2.2 Markowitz Formulation

In our portfolio we have a vector of p stocks. Let the price of a stock at time t be P_t and define the log return as

$$R_{it} = \log(P_{it}) - \log(P_{it-1}). \quad (2.0)$$

The vector of returns for a p set of stocks can then be written as,

$$\mathbf{R}_t = \begin{pmatrix} R_{1t} \\ \vdots \\ R_{pt} \end{pmatrix} \quad (2.1)$$

with the vector of expected returns

$$E(\mathbf{R}_t) = \boldsymbol{\mu} \quad (2.2)$$

and covariance matrix of returns

$$\text{Cov}(R_t) = \boldsymbol{\Sigma} = E(\mathbf{R}_t - \boldsymbol{\mu})(\mathbf{R}_t - \boldsymbol{\mu})'. \quad (2.3)$$

The portfolio weights are w_i for each stock. These must satisfy

$$\sum_{i=1}^p w_i = 1. \quad (2.4)$$

Let w_i be proportion invested in the i -th asset and assuming no short selling, then

$$0 \leq w_i \leq 1. \quad (2.5)$$

The portfolio return P_t at time t is

$$P_t = \sum_{i=1}^p w_i R_{it}. \quad (2.6)$$

The expected return of the portfolio is

$$E(P_t) = \sum_{i=1}^p w_i \mu_i = \mu_p \quad (2.7)$$

and the variance is

$$\begin{aligned} \text{var}(P) &= \mathbf{w} \Sigma \mathbf{w} \\ &= \sum_{i=1}^p \sum_{j=1}^p w_i w_j \sigma_{ij} = \sigma_p^2. \end{aligned} \quad (2.8)$$

with σ_{ij} the ij -th element of Σ

We can change the expected return and variance of the portfolio by changing the weights. We thus need to find an optimal set of weights that gives the best trade-off between expected return and variance.

2.3 The Efficient Frontier

A portfolio is called efficient if it gives the highest possible return for a given level of risk. We can evaluate this using the efficient frontier. The efficient frontier is a 2-dimensional plot of portfolio combinations that have minimum variance, that is, any point on the efficient frontier represents a combination of portfolio weights that give the least possible variance for a given level of return. Equivalently, any point on the efficient frontier gives a portfolio that has the highest expected return for a given level of risk. The efficient frontier plots various minimum variance portfolios with the expected return on the y-axis and risk, measured with the standard deviation, on the x-

axis. Ruppert (2004 pg. 153) points out that if we use the no short selling constraint, we cannot solve the optimization problem using Lagrangian. Thus in order to construct the efficient frontier, we need to solve a constrained quadratic optimization problem. Ruppert (2004 pg. 160) recommends using the portfolio variance as the objective function and the no-short selling rule as one of the linear constraints.

$$\text{Minimize } \sigma_p^2 = \mathbf{w}^T \Sigma \mathbf{w} \quad (2.9)$$

$$\text{such that } \mu_p = \mathbf{w}^T \boldsymbol{\mu}$$

$$\text{and } 0 \leq w_i \leq 1. \quad (2.10)$$

Letting $0 \leq w_i \leq 1$, means that short sales are not considered in this model.

2.4 The Optimal Portfolio

The portfolio optimization algorithm above gives us a set of minimum variance portfolios or *risky* portfolios that a rational investor would choose. In practice, investors do not just hold risky assets, they also hold a risk-free asset in the form of cash or bonds. The investor can even short sale bonds by borrowing in order to invest in the risky portfolio, depending on the investor attitude to risk. The presence of the risk-free asset dramatically changes the set of minimum variance portfolios that a risk-averse investor will actually hold. The efficient frontier that the investor faces after applying the quadratic programming problem in equation (2.9) above is shown as the solid curve in blue in Figure 1.

Now if we take into account the presence of the risk-free asset, the investor can choose to hold a portfolio that contains risky assets from the efficient frontier and the risk-free asset. This portfolio that the investor will hold is shown as the straight line that is tangent to the efficient frontier this is

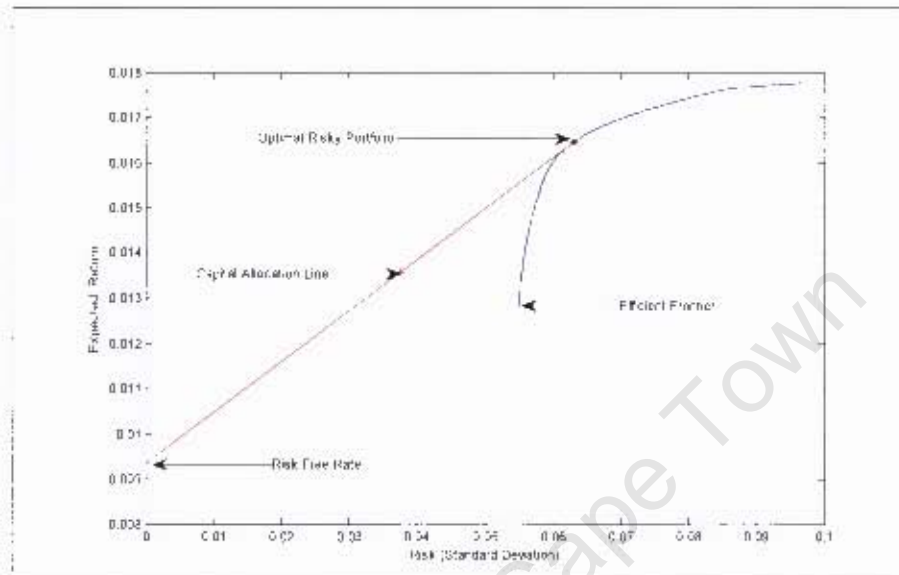


Figure 1: Efficient Frontier and the Optimal Portfolio

termed the Capital Allocation Line (CAL). The CAL intersects the y-axis at the risk-free rate because by definition, the risk-free asset has a standard deviation of zero. Define $E(R_C)$ and σ_C^2 as the expected return and variance of the overall portfolio which includes the risk-free asset, $E(R_A)$ and σ_A^2 as the expected return and variance of the tangent portfolio (the unique risky portfolio that lies on the tangent to the efficient frontier), and R_F as the risk-free rate. Elton and Gruber (1995) show that the equation of the CAL is given by

$$E(R_C) = R_F + \left(\frac{E(R_A) - R_F}{\sigma_A} \right) \sigma_C. \quad (2.11)$$

This is a powerful result because it shows us that the actual portfolio that an investor will hold lies on a straight line that is uniquely determined by the portfolio that has the expected return $E(R_A)$. This portfolio which is called the *optimal portfolio or tangency portfolio*, Ruppert (2004 pg. 143) must be the one that is at the tangent to the efficient frontier. This is because the

tangent portfolio is the one that has the highest reward to risk ratio, shown by the slope term $(E(R_A) - R_F)/\sigma_A$. Ruppert (2004) refers to the slope term as the *Sharpe ratio*, of the portfolio. Any other portfolio other than this optimal portfolio, when combined with the risk-free asset, would have a lower gradient hence it will have a lower reward to risk ratio. Note also that the gradient cannot be greater than $(E(R_A) - R_F)/\sigma_A$ because if it was, it would not even touch the efficient frontier hence it would not be attainable. Any rational investor who is faced with the efficient frontier in Figure 1 will thus have an overall portfolio that consists of the tangent portfolio and the risk free asset, and this overall portfolio lies somewhere on the straight line which is the Capital Allocation Line, (CAL). Of course the precise overall portfolio that the investor will actually hold depends on the degree of risk aversion.

2.5 Empirical Investigation of the Markowitz Model

2.5.1 Objectives and background of the Investigation

In the application of the Markowitz model, our goal is to use empirical data to construct the efficient frontier and also to get an idea of the asset allocation in the optimal portfolio. Using the Markowitz assumptions, we will solve the optimization problem that is shown in equation (2.9). As we build other models later in the thesis, we will frequently compare the results to the Markowitz Model in order to explain the reasons or situations in which Markowitz may or may not be a suitable model for portfolio analysis.

2.5.2 Data used in the investigation

The data we use in the analysis comes from monthly share price data from the Johannesburg Stock Exchange (JSE). This data was obtained from Datas-tream Advance, the online data bank, and also from the JSE Data Bank

Table 1: Shares used in the Portfolio

Share
ABSA GROUP
ANGLO AMERICAN
BARLOWORLD
BIDVEST GROUP
GROUP FIVE
ILLOVO SUGAR
FIRSTRAND
LONMIN (JSE)
MR PRICE GROUP
PICK N PAY HOLDINGS
SHOPRITE
SABMILLER
SAPPI
SASOL
STANDARD BANK

from the Department of Statistical Sciences at the University of Cape Town. The data was then converted into the monthly log returns by differencing the log prices. We decided to use data from January 1996 to February 2008. We needed to choose the time period that we felt was long enough for us to confidently use some asymptotic results regarding the convergence of the estimates to their true quantities. At the same time, too long a time period would not be a true reflection of the current estimates especially since they may change over time. We chose to work with a portfolio consisting of 15 shares which are shown in Table 1. The choice of the number of shares, (15 in our case), becomes a crucial one as we shall see when we discuss the Troskie-Hossain Model, in Chapter 3. A portfolio that has too many shares will be so diversified that it will have very little company specific risk and since this thesis is about estimating risk, we may lose some important insights.

We used the risk-free rate of 11.89% per annum and the proxy for the risk

free rate that we decided to use was the JIBAR (Johannesburg Inter Bank Agreed Rate). This is quoted as a nominal annual compounded quarterly rate (NACQ) so since we are working with monthly log returns, we perform a simple transformation to the JIBAR rate in order to convert it from the annual compounded quarterly rate to the effective monthly log rate.

2.5.3 Method of Investigation

The estimation of the mean and covariance matrix for the analysis was performed in Eviews 3. The constrained quadratic optimization procedure that we need, (shown in equation 2.9), was performed in MATLAB 7. In implementing the procedure, we use the maximum likelihood estimate of the covariance matrix from normal theory which Ruppert (2004 pg. 141) recommends for portfolio analysis. This was used together with the sample mean which we used as the estimate for the vector of expected returns. Assuming that the vector of returns $\mathbf{R}_t \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, for $t = 1, \dots, N$, the maximum likelihood estimates for the mean $\boldsymbol{\mu}$ and variance $\boldsymbol{\Sigma}$, are given by,

$$\hat{\boldsymbol{\mu}} = \bar{\mathbf{R}}_t \text{ and } \hat{\boldsymbol{\Sigma}} = \frac{\sum_{t=1}^N (\mathbf{R}_t - \bar{\mathbf{R}}_t)(\mathbf{R}_t - \bar{\mathbf{R}}_t)'}{N}. \quad (2.12)$$

The risk-free rate was added to the analysis in order to determine the optimal portfolio. The crucial assumption we made in addition to the standard Markowitz assumptions is that the historical estimates of the mean and variance that we calculated from the data are the best estimates of the expected returns and risk that the investor uses.

2.5.4 Results

The efficient frontier obtained from the analysis is shown in Figure 2. The optimal portfolio is marked on the efficient frontier. The investor will thus hold this portfolio and the risk-free asset in proportions that will depend on

Table 2: Composition of the Markowitz Optimal Risky Portfolio

Share	Proportion
Firststrand	6.18%
Lommin	42.69%
Pick n Pay Holdings	0.53%
Shoprite	29.27%
Sasol	21.33%

the attitude to risk. The composition of the optimal portfolio is shown in Table 2. Notice that the optimal portfolio does not include all the original 15 shares. The annualized expected return of the optimal portfolio is 19.8% and the standard deviation is 20.7%.

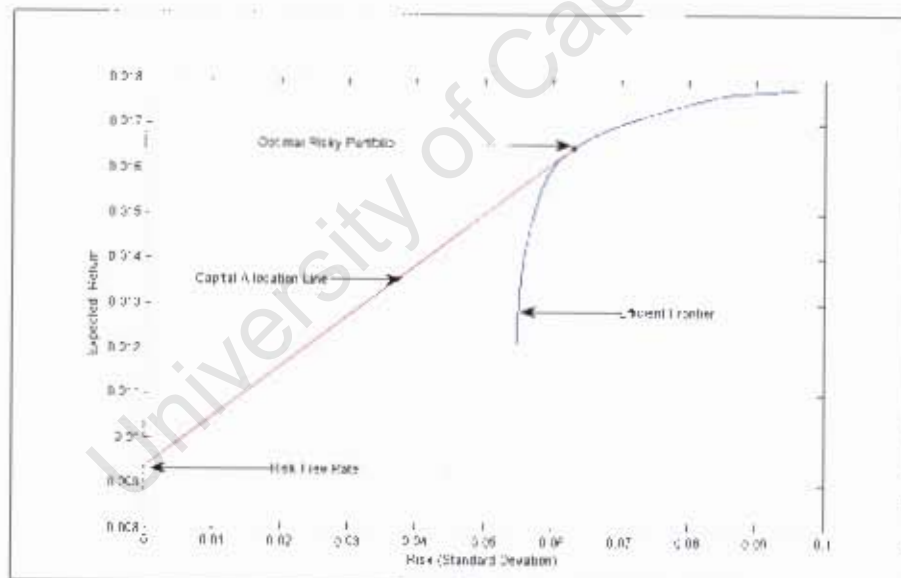


Figure 2: Markowitz Efficient Frontier

2.6 Analysis of the Results and Conclusion

The efficient frontier we obtained gave us the set of portfolio weights that we need in order to construct a minimum variance portfolio. The different points of the efficient frontier are as a result of different weights. The optimal portfolio had a high proportion (about 43%) in a particular stock, Lonmin. Looking at the estimated vector of expected returns and covariance matrix that we used as inputs, we notice that this particular share had the highest expected return and a relatively low standard deviation. This is not surprising since the optimization process penalizes high standard deviations and favours high returns. In addition, further observation of the covariance matrix shows that Lonmin is negatively correlated with Pick n Pay holdings. This is crucial because looking at Equation (2.8), the presence of negative correlation reduces the overall risk of the portfolio due to the negative component of the summation. We will tackle this issue of the presence of negative correlation when we discuss the Troskie-Hossain Model in Chapter 3.

3 Index Models of Asset Returns

The Markowitz model assumes that the asset returns are related to each other through the covariance matrix but does not attempt to explain *how* they are related. If a model could explain exactly how the asset returns from different assets are related then it would be a very useful tool because we would be able to get a better understanding of the underlying market dynamics. This is exactly what William Sharpe (1970) did when he proposed the index models. Before we introduce the Sharpe index models, we first review some key results from the multivariate normal distribution.

3.1 The Bivariate Normal Distribution

The Index models that are used depend on some fundamental results from regression theory. A particular result that is used comes from the properties of the conditional multivariate normal distribution. If $R_t \sim N(\mu_r, \sigma_r^2)$ is the return on a specific share and $I_t \sim N(\mu_I, \sigma_I^2)$ is the return on the overall market such as the JSE All-Share Index or some other market proxy, and the covariance between the market and the share is σ_{rI} , then using matrix notation we can write the following,

$$\begin{pmatrix} R_t \\ I_t \end{pmatrix} \sim N \left[\begin{pmatrix} \mu_r \\ \mu_I \end{pmatrix}, \begin{pmatrix} \sigma_r^2 & \sigma_{rI} \\ \sigma_{Ir} & \sigma_I^2 \end{pmatrix} \right]. \quad (3.1)$$

That means we can write the joint distribution of share and market returns and partition the mean and covariance matrix as shown in Equation (3.1). Using the property of conditional normal random variables it follows that the conditional distribution of $R_t|I_t$ is given by

$$R_t|I_t \sim N \left(\mu_r + \frac{\sigma_{rI}}{\sigma_I^2}(I_t - \mu_I), \sigma_r^2 - \frac{\sigma_{rI}^2}{\sigma_I^2} \right)$$

$$\begin{aligned}
E(R_t|I_t) &= \mu_r + \frac{\sigma_{rI}}{\sigma_I^2}(I_t - \mu_I) & (3.2) \\
&= (\mu_r - \frac{\sigma_{rI}}{\sigma_I^2}\mu_I) + \frac{\sigma_{rI}}{\sigma_I^2}I_t \\
&= \alpha + \beta I_t
\end{aligned}$$

and

$$\begin{aligned}
\text{Var}(R_t|I_t) &= \sigma_r^2 - \frac{\sigma_{rI}^2}{\sigma_I^2} & (3.3) \\
&= \sigma_r^2(1 - \frac{\sigma_{rI}^2}{\sigma_r^2\sigma_I^2}) \\
&= \sigma_r^2(1 - \rho^2).
\end{aligned}$$

This can now be written as

$$R_t|I_t \sim N(\alpha + \beta I_t, \sigma_r^2(1 - \rho^2))$$

3.2 The Sharpe Single Index Model

In the single index model, Sharpe (1970) proposes the following relationship between the return in the i -th share at time t , R_{it} , and the return on some index I_t .

$$R_{it} = \alpha_i + \beta_i I_t + e_{it} \quad (3.4)$$

with

$$E(e_{it}) = 0, \text{ and } E(e_{it}^2) = \sigma_{ei}^2 \quad i = 1, \dots, p \quad (3.5)$$

$$E(e_{it}e_{is}) = 0, \quad t \neq s \quad (3.6)$$

$$E(e_{it}I_t) = 0, \quad t = 1, \dots, N, \quad (3.7)$$

$$E(e_{it}I_t) = 0, \quad t = 1, \dots, N, \quad (3.8)$$

Equation (3.4) states that the return on a share depends on some constant α_i and the product of the return on the overall index I_t and some other constant β_i plus some random noise e_{it} that follows a normal distribution. Equation (3.5) states that the error term has a mean of zero and constant variance. Equation (3.6) shows that the error terms for different time periods are uncorrelated, (no autocorrelation) and Equation (3.7) states that the error terms are uncorrelated with the index. Finally equation (3.8) states that the error terms of different stocks are uncorrelated. This means that the only relationship that individual stocks have between each other is through the market index. This last assumption as shown by Troskie, Guo and Hossain, (2008) is not satisfied in practice, but for the moment we will assume that it holds.

3.3 Sharpe Single Index model in Portfolio Theory

Given the Sharpe Index model assumptions, the portfolio selection problem is solved using the estimated parameters from Equations (3.4) and (3.5). If $E(I_t) = \mu_I$ and $var(I_t) = \sigma_I^2$ is the expected return and variance of the market returns, then the expected return from share i , is given by

$$E_i = E(R_{it}) = \alpha_i + \beta_i \mu_I. \quad (3.9)$$

Note that the time index is not needed since the model assumptions are that α_i , β_i and μ_I are all constant with respect to time. The variance of the returns is given by

$$\begin{aligned} Var(R_{it}) &= var(\alpha_i + \beta_i I_t + e_{it}) \\ &= \beta_i^2 \sigma_I^2 + \sigma_{ei}^2 \\ &= \sigma_{ii}^2 = \sigma_i^2 \end{aligned} \quad (3.10)$$

Equation (3.10) shows that the greater the slope term β_i , the greater the contribution that the market risk σ_I^2 will have on the overall risk. In addition, the overall risk also consists of the risk that is specific to the share $\sigma_{e_i}^2$. The covariance of the returns between different stocks (the off-diagonal elements) are given by

$$\begin{aligned}
 \text{cov}(R_i, R_j) &= E((R_i - E(R_i))(R_j - E(R_j))) \\
 &= E(\beta_i(I - \mu_I) + e_i)(\beta_j(I - \mu_I) + e_j) \\
 &= \beta_i\beta_j\sigma_I^2,
 \end{aligned} \tag{3.11}$$

the last line of equation (3.11) following because $E[\beta_i(I - \mu_I)e_j] = 0$ and $E[e_i e_j] = 0$ for $i \neq j$. The equations above give us the inputs to the portfolio selection problem

$$\begin{aligned}
 &\text{Minimize } \sigma_p^2 = \mathbf{w}^T \Sigma \mathbf{w} \\
 &\text{such that } \mu_p = \mathbf{w}^T \boldsymbol{\mu} \\
 &\text{and } 0 \leq w_i \leq 1.
 \end{aligned} \tag{3.12}$$

where

$$\begin{aligned}
 \mu_p &= \sum_{i=1}^p w_i E_i \\
 &= \sum_{i=1}^p w_i (\alpha_i + \beta_i \mu_I) \\
 &= \left(\sum_{i=1}^p w_i \alpha_i \right) + \left(\sum_{i=1}^p w_i \beta_i \right) \mu_I
 \end{aligned} \tag{3.13}$$

and

$$\begin{aligned}\sigma_p^2 &= \sigma_I^2 \left(\sum_{i=1}^p w_i \beta \right)^2 + \sum_{i=1}^p \sigma_{ei}^2 w_i^2 \\ &= \sum_{i=1}^p \sum_{j=1}^p w_i w_j \sigma_{ij}\end{aligned}$$

with

$$\begin{aligned}\sigma_{ii} &= \sigma_i^2 \\ &= \beta_i^2 \sigma_I^2 + \sigma_{ei}^2 \text{ and} \\ \sigma_{ij} &= \beta_i \beta_j \sigma_I^2.\end{aligned}$$

or more generally, σ_{ij} is an element from the following covariance matrix,

$$\text{cov}(R) = \sigma_I^2 \boldsymbol{\beta} \boldsymbol{\beta}' + \begin{pmatrix} \sigma_{e1}^2 & 0 & \dots & 0 \\ 0 & \sigma_{e2}^2 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & \dots & \sigma_{ep}^2 \end{pmatrix}$$

The single index model assumes that we know the values of α_i , β_i and σ_{ei}^2 . Rupert (2004 pg. 174) recommends using the least squares estimates. If the model assumptions are not seriously violated, then the least squares estimates are the best. In fact to be more precise, the Gauss-Markoff Theorem tells us that the least square estimates are the best linear unbiased estimators. If however these assumptions are violated, then ordinary least squares estimates give very unreliable estimates. In later chapters we will use other estimating techniques that are more robust compared to the least squares estimates when the assumptions are violated.

3.4 Empirical Investigation-Sharpe Single Index Model

3.4.1 Objectives of the Investigation

Using the Sharpe index model, we would like to estimate the parameters α_i , β_i and σ_{ei}^2 using historical data from the Johannesburg Stock Exchange. These estimates will then be used in order to estimate the mean and covariance matrix of the asset returns which we will then use as inputs in the optimal portfolio selection algorithm. We would also like to compare the efficient frontier that we get using the Sharpe single index model to that of the Markowitz model which we found in Chapter 2. We will be interested in comparing the changes in the composition of the optimal risky portfolio when we move from the Markowitz model to the Sharpe model.

3.4.2 Data used in the Investigation

For consistency and comparison purposes, we will use the same share data we used in Chapter 2, (see Table 1), with the only addition being the index on the JSE AllShare Index over the same time period.

3.4.3 Method of Investigation

The data were analyzed in Eviews3 and the parameters were estimated using least squares. Eviews3 gives us information on the parameter estimates, standard errors and the overall fit of the model. The estimates that Eviews3 gives for the standard errors are the unbiased estimates. Using the estimated parameters, the mean and covariance matrix were calculated and these were then used in Matlab7 in the optimization process. Matlab gives the efficient frontier and details on the composition of the optimal portfolio. The vector of expected returns and the covariance matrix were calculated from Equation

Table 3: Sharpe Single Index Model Estimates

Share	$\hat{\alpha}_i$	$\hat{\beta}_i$	β_i t-stat	β_i stderr	$\hat{\sigma}_{ei}$	adjusted-R ²
Absa	-0.00036	1.078	9.78	0.110	0.082	0.40
Anglos	-0.00064	1.305	16.70	0.078	0.058	0.66
Barloword	-0.00281	1.001	10.87	0.093	0.069	0.45
Bidvest	0.0009	0.784	9.77	0.080	0.059	0.40
Group 5	0.00281	0.726	4.18	0.174	0.128	0.10
Illovo	0.00026	0.727	6.91	0.105	0.078	0.24
Firstrand	0.00279	1.025	9.73	0.105	0.078	0.39
Lonmin	0.00814	0.872	7.97	0.110	0.081	0.30
Mr price	-0.00276	0.944	5.93	0.159	0.118	0.19
Pick n pay	0.00569	0.624	5.32	0.117	0.087	0.16
Shoprite	0.00784	0.600	5.09	0.118	0.087	0.15
SAB	-0.00344	0.866	12.20	0.071	0.053	0.51
Sappi	-0.00791	1.158	9.37	0.124	0.092	0.38
Sasol	0.0055	1.069	9.86	0.108	0.080	0.40
Standard B	0.00131	0.942	10.48	0.090	0.067	0.43

(3.13) and these were used to solve the quadratic optimization problem in Equation (3.12).

3.4.4 Results

Table 3 shows the various estimated coefficients. The beta or slope coefficients are very important statistics because they are crucial inputs into the estimation of portfolio risk. In addition to getting the parameter estimates, we would also like to know how reliable these estimates are. Keeping in mind that our results are only estimates of the true unknown values, it is good practice in any statistical estimation work to always give the standard errors of the estimates. The adjusted R-Squared statistic has also been included which gives us a rough indication of the proportion of variation that is explained by the market.

Figure 3 shows the efficient frontier produced from the Sharpe single index

model. For comparison, the corresponding efficient frontier obtained from Markowitz is plotted on the same graph. The green line shows the Sharpe while the blue represents the Markowitz model. The optimal risky portfolio is also shown on the curves. The graph clearly shows that for most of the levels of expected return, the Sharpe model has a higher risk since it is shifted to the right. Notice also that although the Sharpe optimal portfolio has a slightly higher return, it has a significantly higher risk. That means that either the Sharpe model is overestimating the risk or the Markowitz model is underestimating the risk. As it stands, it looks as if rational investor would prefer the Markowitz model because it is more efficient.

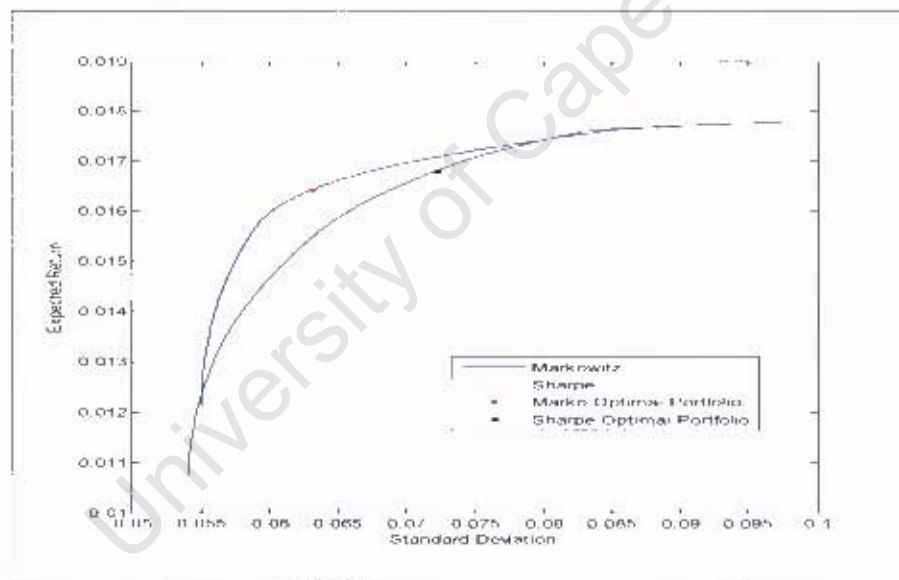


Figure 3: Sharpe Single Index Efficient Frontier

Table 4 compares the composition of the optimal risky portfolio between the Sharpe and the Markowitz models. The composition and weights of the shares are quite different. These differences will have a big impact in portfolio performance. It is thus very important to try and understand the reasons for

Table 4: Composition of the Markowitz and Sharpe Optimal Portfolios

Share	Markowitz	Sharpe
Firststrand	6.18%	0%
Lonmin	42.69%	45.51%
Pick n Pay Holdings	0.53%	2.46%
Shoprite	29.27%	20.83%
Sasol	21.33%	31.2%

these differences so that, depending on the situation, the practitioner will be able to decide which of the two models is appropriate.

3.4.5 Conclusion

The results from the Sharpe model are quite different to those of the Markowitz model in that they suggest that the Markowitz model is more efficient than the Sharpe model. Does this mean that Markowitz is better? Probably not, because we still need to get a better understanding of how the covariance structure of the returns affect this. In order to investigate this issue, we now present the Troskie-Hossain Improved Sharpe model proposed by Hossain, Troskie and Guo, (2005) in the following section.

3.5 The Troskie-Hossain Improved Sharpe Model

The Troskie-Hossain Improved Sharpe Model is an extension of the Sharpe Index Model with the relaxation of a crucial assumption regarding the correlation of the residuals. Under the Sharpe model, the only relationship different shares had between each other was through the market index. If the error terms are correlated, then it means that there is some correlation between the share returns that cannot be explained by the market. Recall

that the Sharpe model was formulated as,

$$R_{it} = \alpha_i + \beta_i I_t + e_{it} \quad (3.15)$$

with

$$E(e_{it}) = 0, \text{ and } E(e_{it}^2) = \sigma_{ei}^2 \quad i = 1, \dots, p \quad (3.16)$$

$$E(e_{it}e_{is}) = 0, \quad t \neq s \quad (3.17)$$

$$E(e_{it}I_t) = 0, \quad t = 1, \dots, N, \quad (3.18)$$

$$E(e_{it}I_t) = 0, \quad t = 1, \dots, N. \quad (3.19)$$

This implies that

$$\text{cov}(R) = \sigma_j^2 \beta \beta' + \begin{pmatrix} \sigma_{e1}^2 & 0 & \dots & 0 \\ 0 & \sigma_{e2}^2 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & \dots & \sigma_{ep}^2 \end{pmatrix}. \quad (3.20)$$

Sharpe assumes that all e_{it} , $i = 1, \dots, p$, $t = 1, \dots, N$ are uncorrelated. In the Troskie-Hossain Improved Sharpe model, this assumption is relaxed by assuming that the disturbances e_{it} of the different stocks are dependent (correlated). Troskie, Guo and Hossain (2008) have found evidence of this correlation on the JSE shares.

Thus

$$\begin{aligned} E(e_{it}e_{jt}) &= \sigma_{ij}, \quad i \neq j \\ &= \sigma_{ei}^2 \quad i = j \end{aligned} \quad (3.21)$$

or

$$E(\mathbf{ee}') = \mathbf{\Omega} = \begin{pmatrix} \sigma_{e1}^2 & \sigma_{12} & \dots & \sigma_{1p} \\ \sigma_{21} & \sigma_{e2}^2 & \dots & \sigma_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ \sigma_{p1} & \dots & \dots & \sigma_{ep}^2 \end{pmatrix} \quad (3.22)$$

with

$$\text{cov}(\mathbf{R}) = \sigma_I^2 \boldsymbol{\beta} \boldsymbol{\beta}' + \mathbf{\Omega} = \mathbf{\Phi} \quad (3.23)$$

For portfolio $P = \mathbf{w}'\mathbf{R}$ we have

$$E(P) = \mathbf{w}'(\boldsymbol{\alpha} + \boldsymbol{\beta}\mu_I) = \mu_p \quad (3.24)$$

and

$$\begin{aligned} \text{var}(P) &= \mathbf{w}'(\sigma_I^2 \boldsymbol{\beta} \boldsymbol{\beta}' + \mathbf{\Omega})\mathbf{w} \\ &= \mathbf{w}'\mathbf{\Phi}\mathbf{w} = \sigma^2. \end{aligned} \quad (3.25)$$

These inputs to portfolio selection can then be used in Equation (3.12) to solve the quadratic programming problem. To estimate $\mathbf{\Omega}$, let

$$\hat{\mathbf{E}} = \begin{pmatrix} \hat{e}_{11} & \hat{e}_{12} & \dots & \hat{e}_{1N} \\ \hat{e}_{21} & \hat{e}_{22} & \dots & \hat{e}_{2N} \\ \vdots & \vdots & \vdots & \vdots \\ \hat{e}_{p1} & \dots & \dots & \hat{e}_{pN} \end{pmatrix} \quad (3.26)$$

$$\text{with } \hat{e}_i = R_{it} - E(R_{it}). \quad (3.27)$$

Troskie, Guo and Hossain, (2008) recommend using the unbiased estimator

$$\hat{\Omega} = \frac{1}{N-2} \hat{\mathbf{E}}\hat{\mathbf{E}}'. \quad (3.28)$$

and

$$\hat{\Phi} = \hat{\sigma}_I^2 \hat{\beta} \hat{\beta}' + \hat{\Omega}. \quad (3.29)$$

This means that the Troskie-Hossain model will be different to Sharpe depending on the magnitude of the off-diagonal elements of $\hat{\Omega}$.

3.6 Empirical Investigation of the Troskie-Hossain Improved Sharpe Single Index Model

3.6.1 Objectives of the Investigation

The aims of the investigation are threefold. Firstly, we would like to compare the efficient frontiers produced from the Troskie-Hossain model to the Sharpe model. Secondly, if there are any differences then we would like to use the covariance matrix of the residuals in order to explain these differences because the only difference between the models is due to the off-diagonal elements of the residual covariance matrix. Thirdly, if we can explain these differences, then we would like to determine what impact will this have for a portfolio manager.

3.6.2 Methodology

The method used is identical to the one used in the Sharpe model. The only difference is that we use the covariance matrix from the Troskie-Hossain model instead of the one from the Sharpe model.

3.6.3 Results

Table 5 shows the risk return properties of the Markowitz and Troskie-Hossain models. Figure 4 shows the Troskie-Hossain efficient frontier in green and the Sharpe model in red. For comparison, the Markowitz model is also shown in blue. Note that the Markowitz model is almost identical to the Troskie-Hossain model, so identical in fact, that the efficient frontiers are difficult to distinguish. Does that mean that we should not bother to use Troskie-Hossain and just rather use Markowitz? No! The beauty of the Troskie-Hossain model is that it can easily be generalized to include time-series models such as GARCH and ARMA models. The Troskie-Hossain model has thus resulted in a left-shifting of the efficient frontier compared to Sharpe. This is good news to the portfolio manager but we need to understand exactly why this has happened and for that we turn to the covariance matrix of the residuals.

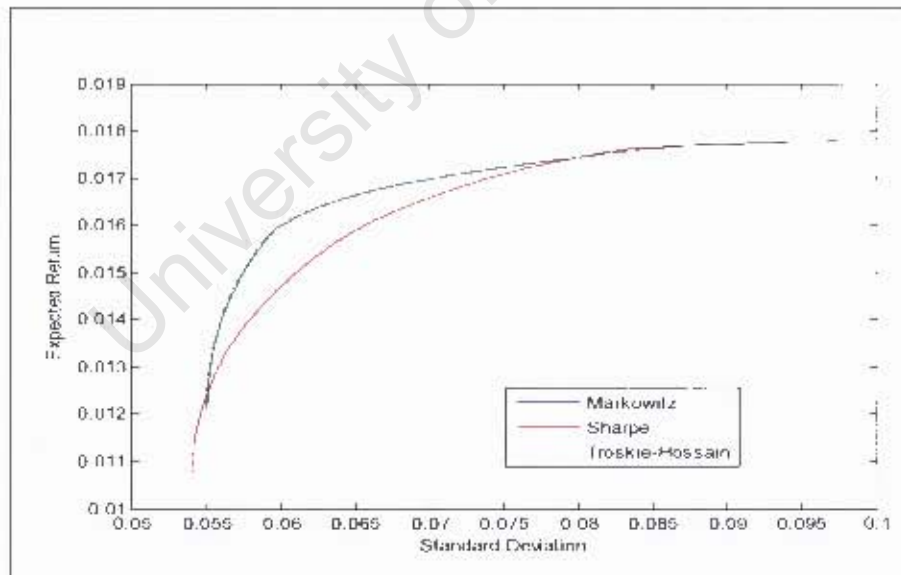


Figure 4: Troskie-Hossain and Sharpe Efficient Frontiers

Table 5: Risk-Return properties of the optimal portfolio

Method	Expected Return	Standard Deviation
Markowitz	19.8%	21.89%
Sharpe	20.16%	25.01%
Tros-Hos	19.68%	21.93%

Figure 5: Single Index Residual Correlation Matrix

1.00	-0.36	0.07	0.27	0.03	-0.03	0.49	-0.21	0.18	0.26	0.19	0.10	-0.18	-0.30	0.56
-0.36	1.00	-0.19	-0.35	-0.01	0.04	-0.44	0.31	-0.26	-0.22	-0.27	-0.14	0.24	0.33	-0.53
0.07	-0.19	1.00	0.42	0.31	-0.18	0.14	-0.22	0.17	0.42	0.37	0.13	0.09	-0.11	0.00
0.27	-0.35	0.42	1.00	0.22	-0.01	0.47	-0.31	0.23	0.39	0.42	0.13	-0.08	-0.32	0.26
0.03	-0.01	0.31	0.22	1.00	-0.03	0.07	-0.14	0.05	0.27	0.11	0.18	0.10	0.01	0.05
-0.03	0.04	-0.18	-0.01	-0.03	1.00	0.00	0.11	0.02	-0.27	0.07	-0.08	0.00	0.04	-0.05
0.49	-0.44	0.14	0.47	0.07	0.00	1.00	-0.45	0.05	0.34	0.35	0.14	-0.18	-0.36	0.51
-0.21	0.31	-0.22	-0.31	-0.14	0.11	-0.45	1.00	-0.24	-0.31	-0.37	-0.13	0.08	0.08	-0.29
0.18	-0.26	0.17	0.23	0.05	0.02	0.05	-0.24	1.00	0.15	0.20	-0.07	-0.12	-0.20	0.06
0.26	-0.22	0.42	0.39	0.27	-0.27	0.34	-0.31	0.15	1.00	0.39	0.28	0.10	-0.21	0.33
0.19	-0.27	0.37	0.42	0.11	0.07	0.35	-0.37	0.20	0.39	1.00	0.10	-0.07	-0.20	0.15
0.10	-0.14	0.13	0.13	0.18	-0.08	0.14	-0.13	-0.07	0.28	0.10	1.00	-0.08	-0.17	0.21
-0.18	0.24	0.09	-0.08	0.10	0.00	-0.18	0.08	-0.12	0.10	-0.07	-0.08	1.00	0.32	-0.09
-0.30	0.33	-0.11	-0.32	0.01	0.04	-0.36	0.08	-0.20	-0.21	-0.20	-0.17	0.32	1.00	-0.17
0.56	-0.53	0.00	0.26	0.05	-0.05	0.51	-0.29	0.06	0.33	0.15	0.21	-0.09	-0.17	1.00

The correlation matrix of the residuals shown in Table 5, shows the presence of negative and positive correlations. This is a problem because Sharpe assumes that these are zero. We can actually do a formal statistical test in order to test for the significant correlations. We can use the distributional properties of the sample correlation to test for significance as follows:

The null hypothesis is

$$H_0 : \rho = 0,$$

the alternative is

$$H_1 : \rho \neq 0.$$

The test statistic T is given by

$$T = r \sqrt{\frac{n-2}{1-r^2}} \sim t_{n-2}.$$

Using a multiple t-test with the Bonferroni corrected significance level, the null hypothesis is rejected for most of the correlations. In order to investigate how the presence of these correlation affect portfolio risk we observe that the total portfolio risk can be written as,

$$\begin{aligned} \text{var}(P) &= \mathbf{w}'(\sigma_I^2 \boldsymbol{\beta} \boldsymbol{\beta}' + \boldsymbol{\Omega}) \mathbf{w} \\ &= \mathbf{w}' \sigma_I^2 \boldsymbol{\beta} \boldsymbol{\beta}' \mathbf{w} + \mathbf{w}' \boldsymbol{\Omega} \mathbf{w} \end{aligned} \quad (3.30)$$

Notice that it is the second term, $\mathbf{w}' \boldsymbol{\Omega} \mathbf{w}$, in Equation (3.30) that differentiates Sharpe from Troskie-Hossain. Haines, (2008) pointed out that the $\mathbf{W}' \boldsymbol{\Omega} \mathbf{W}$ term is very important in portfolio risk. This term can be written as

$$\begin{aligned} \mathbf{w}' \boldsymbol{\Omega} \mathbf{w} &= \sum_{i,j}^p w_i w_j \sigma_{i,j} \\ &= \sum_i^p w_i^2 \sigma_i^2 + \sum_{i \neq j}^p w_i w_j \sigma_{ij} \end{aligned} \quad (3.31)$$

Sharpe assumes that $\sum_{i \neq j}^p w_i w_j \sigma_{ij} = 0$, while Troskie, Hossain and Guo do not. If there are positive covariances, Sharpe underestimates portfolio risk and if there are negative covariances, Sharpe overestimates it. That is a heuristic explanation because it does not explain what would happen if there are both negative and positive correlations. Negative covariances are very good because they reduce the portfolio risk. Keep in mind that the optimization process is trying to minimize portfolio risk so it will be biased to selecting shares with these negative covariance because they will result in the lowest risk. From the particular correlation matrix that was obtained in our em-

irical study, we found more negative correlations than positive correlations. Sharpe did not pick these negative correlations up and so this resulted in higher portfolio risk. This explains the right-shift of the efficient frontier. How does this impact on a portfolio manager? Well, recall that using the Sharpe model we can decompose the total portfolio risk into two parts as,

$$\text{Var}(P) = \sigma_I^2(\boldsymbol{\beta}'\mathbf{w})^2 + \mathbf{w}'\boldsymbol{\Omega}\mathbf{w} \quad (3.32)$$

The first term, $\sigma_I^2(\boldsymbol{\beta}'\mathbf{w})^2$ is called the market or systematic risk and the second term $\mathbf{w}'\boldsymbol{\Omega}\mathbf{w}$ is the stock-specific or unsystematic risk. A portfolio manager, no matter how skillful, cannot get rid of the systematic risk. The benefits of portfolio management result from the fact that the manager can add value by reducing the unsystematic risk. This confirms the importance of estimating the $\mathbf{w}'\boldsymbol{\Omega}\mathbf{w}$ term properly. What happens as the number of shares in our portfolio becomes arbitrarily large? Assuming the Sharpe model is correct and that we have N shares in an equally weighted portfolio, Elton and Gruber, (1995) showed quite easily that the specific risk term $\mathbf{w}'\boldsymbol{\Omega}\mathbf{w} \rightarrow 0$ as $N \rightarrow \infty$. What if the Sharpe model is not appropriate, specifically what happens when the off-diagonal elements of the residual covariance matrix are not zero? Then it can be shown that in this case we have

$$\mathbf{w}'\boldsymbol{\Omega}\mathbf{w} \rightarrow \bar{\sigma}_{ij} \text{ as } N \rightarrow \infty. \quad (3.33)$$

So with the Troskie-Hossain model, we cannot simply ignore these covariances no matter how large the portfolio is. Notice that because $\mathbf{w}'\boldsymbol{\Omega}\mathbf{w}$ converges to the average covariance $\bar{\sigma}_{ij}$, it may well be possible that there are both large negative and positive covariances, meaning that $\bar{\sigma}_{ij}$ is near zero anyway which would produce similar results to Sharpe. If however the majority of the covariances are negative, then this will result in a substantial difference in the estimation of risk and it will cause a significant left shift of the efficient frontier. In any case, it is quite evident that Sharpe can lead

to an inaccurate measure of portfolio risk. We need to keep in mind, as Troskie, Guo and Hossain, (2008) point out, that portfolios are not equally weighted so the contribution of a few large covariances in the residual covariance matrix can make a big difference. That means the estimate Troskie Guo and Hossain (2008) used, $\mathbf{w}'\Omega\mathbf{w} \rightarrow \bar{\sigma}_{ij}$ as $N \rightarrow \infty$, from an equally weighted portfolio may actually be an underestimation of risk if we drop the assumption that the weights are equal.

3.6.4 Conclusion

The study of the Sharpe index model as well as the Troskie-Hossain Improved Sharpe Model has provided valuable insights into portfolio construction which we can summarize as follows:

- The efficient frontier and the optimal portfolio produced from the Sharpe Model can be very different to those produced by Markowitz.
- If the assumption that the residuals between different stocks are uncorrelated is dropped and we use the Troskie-Hossain model, then we can get a very different looking efficient frontier when compared to the Sharpe model.
- The Troskie-Hossain model produces an efficient frontier that is very similar to Markowitz. However Troskie-Hossain may be preferred because it is quite easy to generalize it to include time series issues such as autocorrelation and heteroskedasticity in the residuals. This is because if the residuals are autocorrelated or heteroskedastic, we can use AR or GARCH terms in the least squares estimation.
- The Sharpe model can overestimate the portfolio risk when there are negative covariances in the residual covariance matrix and underestimate the risk when they are positive covariances.

- Since the portfolio optimization process is an algorithm that minimizes portfolio risk, the presence of negative covariance terms in particular, will have a big impact since the optimizer will tend to favor these lower covariances.
- Even when the portfolio size becomes arbitrarily large, we simple cannot ignore the $\mathbf{w}'\Omega\mathbf{w}$ term because unlike in the Sharpe model, Troskie et al., (2008) showed that this term does not always vanish.

3.7 The Sharpe Multiple Index Models

The Sharpe and the Troskie-Hossain models of asset returns model the returns on the share as a function of a single market index as

$$R_{it} = \alpha_i + \beta_i I_t + e_{it}. \quad (3.34)$$

The random error e_{it} term represents any other variation in the returns R_{it} that cannot be explained by the index I_t . There may well be other indices or fundamental variables that can explain movements that the market index cannot. This would be very useful because some of the variation that was previously unexplained by the single index model can now be attributed to another factor meaning that the residual variance can be reduced. From a financial modeling perspective, this reduction in the variance is very important because it could improve the ability of the model to forecast. More formally, the multiple index model can be represented as

$$\begin{aligned} R_{it} &= \alpha_i + \beta_{i1}I_1 + \beta_{i2}I_2 + \dots + \beta_{iM}I_M + e_{it}, \\ i &= 1, \dots, p, \quad t = 1, \dots, N, \end{aligned} \quad (3.35)$$

with each I_i representing the return on an index with the following assumptions

$$E(e_{it}^2) = \sigma_{ei}^2 \quad (3.36)$$

$$E(e_{it}e_{is}) = 0, \quad t \neq s = 1, \dots, N, \quad (3.37)$$

$$E(e_{it}I_{jt}) = 0, \quad j = 1, \dots, M, \quad t = 1, \dots, N, \quad (3.38)$$

$$E(e_{it}e_{jt}) = 0, \quad t = 1, \dots, N. \quad (3.39)$$

$$E(I_{jt}I_{kt}) = c_{jk}, \quad j, k = 1, \dots, M \quad (3.40)$$

Note that as for the Sharpe model, the assumptions regarding lack of correlation of residuals are still used and the only new assumption we need is regarding the correlation of the indices. Equation (3.40) shows that the indices that are used are assumed to be correlated with each other. The expected return on each share is now given by

$$\begin{aligned} E_i &= E(R_i) = \alpha_i + \beta_{i1}\mu_1 + \dots + \beta_{iM}\mu_M, \\ i &= 1, \dots, p. \end{aligned} \quad (3.41)$$

This can be written more clearly by introducing

$$\boldsymbol{\alpha} = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_p \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_{11} & \cdots & \beta_{1M} \\ \vdots & \vdots & \vdots \\ \beta_{p1} & \cdots & \beta_{pM} \end{pmatrix}, \quad \mathbf{e}_t = \begin{pmatrix} e_{1t} \\ \vdots \\ e_{pt} \end{pmatrix}$$

$$\mathbf{R}_t = \begin{pmatrix} R_{1t} \\ \vdots \\ R_{pt} \end{pmatrix} \text{ and } \mathbf{I}_t = \begin{pmatrix} I_{1t} \\ \vdots \\ I_{Mt} \end{pmatrix}, \quad t = 1, \dots, N.$$

Then

$$\mathbf{R}_t = \boldsymbol{\alpha} + \boldsymbol{\beta}\mathbf{I}_t + \mathbf{e}_t \quad (3.42)$$

with

$$E(\mathbf{R}_t) = \boldsymbol{\alpha} + \boldsymbol{\beta}\boldsymbol{\mu}_1 \quad (3.43)$$

and

$$\boldsymbol{\mu}_1 = E(\mathbf{I}_t) = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_M \end{pmatrix}. \quad (3.44)$$

The covariance matrix of \mathbf{R}_t is given by

$$\begin{aligned} \text{cov}(\mathbf{R}_t) &= E[\mathbf{R}_t - E(\mathbf{R}_t)][\mathbf{R}_t - E(\mathbf{R}_t)]' \\ &= E[\boldsymbol{\alpha} + \boldsymbol{\beta}\mathbf{I}_t + \mathbf{e}_t - \boldsymbol{\alpha} - \boldsymbol{\beta}\boldsymbol{\mu}_1][\boldsymbol{\alpha} + \boldsymbol{\beta}\mathbf{I}_t + \mathbf{e}_t - \boldsymbol{\alpha} - \boldsymbol{\beta}\boldsymbol{\mu}_1]' \\ &= E[\boldsymbol{\beta}(\mathbf{I}_t - \boldsymbol{\mu}_1) + \mathbf{e}_t][\boldsymbol{\beta}(\mathbf{I}_t - \boldsymbol{\mu}_1) + \mathbf{e}_t]' \\ &= \boldsymbol{\beta}E(\mathbf{I}_t - \boldsymbol{\mu}_1)(\mathbf{I}_t - \boldsymbol{\mu}_1)'\boldsymbol{\beta}' + E(\mathbf{e}_t\mathbf{e}_t') \quad \text{since } E(\mathbf{I}_t\mathbf{e}_t') = 0 \\ &= \boldsymbol{\beta}\mathbf{C}\boldsymbol{\beta}' + \boldsymbol{\Omega} \\ &= \boldsymbol{\Phi}. \end{aligned} \quad (3.45)$$

Note that using the assumptions of Sharpe, the matrix $\boldsymbol{\Omega} = E(\mathbf{e}_t\mathbf{e}_t')$ is a diagonal matrix since Sharpe assumes the residuals are uncorrelated. Our primary objective is to now use the multiple index model in the portfolio allocation problem. If \mathbf{w} is the vector of portfolio weights and E_p and σ_p^2 represent the expected return and variance of the portfolio, then we have

$$E_p = E(P) = \mathbf{w}'(\boldsymbol{\alpha} + \boldsymbol{\beta}\boldsymbol{\mu}_1) \quad (3.46)$$

and

$$\sigma_p^2 = \text{var}(P) = \mathbf{w}'\boldsymbol{\Phi}\mathbf{w}. \quad (3.47)$$

The portfolio problem is then

$$\text{Minimize } \sigma_p^2 = \mathbf{w}^T \Phi \mathbf{w} \quad (3.48)$$

$$\text{such that } \mu_p = \mathbf{w}^T \boldsymbol{\mu} \quad (3.49)$$

$$\text{and } 0 \leq w_i \leq 1.$$

3.8 The Troskie-Hossain Improved Multiple Index Model

The Troskie-Hossain improved multiple index model modifies the Sharpe multiple index model by relaxing the assumption that the residuals between different shares are uncorrelated. More formally, the Troskie-Hossain model can be represented as.

$$\begin{aligned} R_{it} &= \alpha_i + \beta_{i1}I_1 + \beta_{i2}I_2 + \dots + \beta_{iM}I_M + e_{it}, \\ i &= 1, \dots, p, t = 1, \dots, N, \end{aligned} \quad (3.50)$$

with the following assumptions

$$E(e_{it}^2) = \sigma_{ei}^2 \quad (3.51)$$

$$E(e_{it}e_{is}) = 0, t \neq s = 1, \dots, N, \quad (3.52)$$

$$E(e_{it}I_{jt}) = 0, j = 1, \dots, M, t = 1, \dots, N, \quad (3.53)$$

$$E(e_{it}e_{jt}) = 0, t = 1, \dots, N. \quad (3.54)$$

$$E(I_{jt}I_{kt}) = c_{jk}, j, k = 1, \dots, M. \quad (3.55)$$

Using the matrix representation similar to the one used for Sharpe above we now have

$$\mathbf{R}_t = \boldsymbol{\alpha} + \boldsymbol{\beta}\mathbf{I}_t + \mathbf{e}_t \quad (3.56)$$

with

$$E(\mathbf{R}_t) = \boldsymbol{\alpha} + \boldsymbol{\beta}\boldsymbol{\mu}_1 \quad (3.57)$$

and

$$\boldsymbol{\mu}_1 = E(\mathbf{I}_t) = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_M \end{pmatrix}. \quad (3.58)$$

The covariance matrix of \mathbf{R}_t is given by

$$\begin{aligned} cov(\mathbf{R}_t) &= E[\mathbf{R}_t - E(\mathbf{R}_t)][\mathbf{R}_t - E(\mathbf{R}_t)]' \\ &= E[\boldsymbol{\alpha} + \boldsymbol{\beta}\mathbf{I}_t + \mathbf{e}_t - \boldsymbol{\alpha} - \boldsymbol{\beta}\boldsymbol{\mu}_1][\boldsymbol{\alpha} + \boldsymbol{\beta}\mathbf{I}_t + \mathbf{e}_t - \boldsymbol{\alpha} - \boldsymbol{\beta}\boldsymbol{\mu}_1]' \\ &= E[\boldsymbol{\beta}(\mathbf{I}_t - \boldsymbol{\mu}_1) + \mathbf{e}_t][\boldsymbol{\beta}(\mathbf{I}_t - \boldsymbol{\mu}_1) + \mathbf{e}_t]' \\ &= \boldsymbol{\beta}E(\mathbf{I}_t - \boldsymbol{\mu}_1)(\mathbf{I}_t - \boldsymbol{\mu}_1)'\boldsymbol{\beta}' + E(\mathbf{e}_t\mathbf{e}_t') \quad \text{since } E(\mathbf{I}_t\mathbf{e}_t') = 0 \\ &= \boldsymbol{\beta}\mathbf{C}\boldsymbol{\beta}' + \boldsymbol{\Omega}_{tros} \\ &= \boldsymbol{\Phi}. \end{aligned} \quad (3.59)$$

The matrix $\boldsymbol{\Omega}_{tros} = E(\mathbf{e}_t\mathbf{e}_t')$ under the Troskie-Hossain model is no longer diagonal. We need to be very careful when we estimate this matrix because unlike the single index model, the multiple index model now contains a large number of parameters. Using the incorrect degrees of freedom can make a substantial difference particularly when time-series models are used. Depending on the type of work, one may prefer to use either the consistent estimators or the unbiased estimators. For our purposes we prefer to use the unbiased estimators. To estimate $\boldsymbol{\Omega}_{tros}$ from a data set of N points and a model with $M + 1$ (including the constant term α) parameters, Troskie, Hossain and Guo (2008) recommend using,

$$\hat{\boldsymbol{\Omega}}_{tros} = \frac{1}{N - M - 1} \hat{\mathbf{E}}\hat{\mathbf{E}}'. \quad (3.60)$$

3.9 Empirical Investigation of the Multiple Index Models

3.9.1 Objectives of the Investigation

We would like to investigate the effect of using multiple index models on the efficient frontier and on the composition of the optimal portfolio. The Sharpe multiple index model will be compared to the Sharpe single index model in order to investigate the differences that result from using multiple explanatory variables instead of a single index. Particular attention will be focused to how the level of risk of the optimal portfolio changes depending on the model used. The Sharpe multiple index model will be compared to the Troskie-Hossain multiple index model in order to determine the differences in the resulting efficient frontier, particularly with respect to risk.

3.9.2 Data used in the investigation

Table 6 shows the list of indices that were chosen initially. The indices were obtained from the Datastream Database and they were converted into log return series to be used as the explanatory variables in the multiple index models.

3.9.3 Methodology

With the multiple index models we are now faced with an additional problem of actually choosing the explanatory variables that we will use in our model. Unlike the single index model that assumes that the market index is the only explanatory variable, we now have to decide not only which variables to use but also how many variables to use. This is not an easy process but this is certainly an area that an experienced practitioner can add value in terms

Table 6: Explanatory Variables

Index	Variable
Jse Banks	x1
Jse Financials	x2
Jse Financial & Industrials	x3
Jse Gold Miners	x4
Jse Industrials	x5
Jse Mining	x6
Jse All Share	x7
Jse Platinum	x8
Jse excl. Resources	x9
Rand USD Exchange rate	x10

of identifying the explanatory variables. In the investigation carried out, we start by choosing ten indices as possible explanatory variables. The indices were selected on the basis of their macroeconomic relationships with the underlying shares. We are not sure which of these variables have a significant relationship with the shares so we will use a backward stepwise elimination procedure. In this model selection algorithm, we start by fitting a regression model with all 10 explanatory variables. We then look at the t-statistics in order to decide which variables can be eliminated. A t-statistic with an absolute value of less than 1.96 (the 5% significance level) is considered to be insignificant. We however do not start by eliminating all the insignificant variable at once, rather, we remove the least significant one first and then refit the models with the remaining variables. This process of eliminating insignificant variables continues until we have 5 explanatory variables for each share. Note that different shares will have different explanatory variables but our aim is to get 3 or 4 common indices that will give a reasonable fit for all the shares. At this stage we make use of the adjusted R^2 and the Schwartz Criterion, which are both ways of evaluating different models. We favour models with the highest adjusted R^2 statistic and models with the lowest Schwartz Criterion statistic.

3.9.4 Results

The model selection procedure indicated that there were 3 explanatory variables that explained most of the variation in all the 15 shares. These were the Jse Gold Index (x4), Jse All Share index (x7) and the Rand/USDollar exchange rate (x10). The associated regression coefficients β_{gold} , β_{jse} , and β_{exch} together with the t-statistics are shown in Figure 6.

Figure 6: Multiple Index Model Betas

Share	β_{gold}	t-stat	β_{jse}	t-stat	β_{exch}	t-stat
Absa	-0.26	-4.05	1.29	11.06	-0.45	-3.17
Anglos	0.10	2.21	1.23	14.57	0.38	3.70
Barlword	-0.08	-1.38	1.06	10.36	-0.43	-3.51
Bidvest	-0.22	-4.70	0.96	11.49	-0.29	-2.87
Group 5	-0.04	-0.34	0.74	3.73	-0.35	-1.45
Illovo	0.08	1.16	0.67	5.53	0.15	1.05
Firstrand	-0.25	-4.06	1.22	11.03	-0.45	-3.32
Lonmin	0.18	2.72	0.75	6.41	0.58	4.09
Mr price	-0.15	-1.51	1.05	5.88	-0.53	-2.44
Pick n pay	-0.18	-2.49	0.76	5.88	-0.42	-2.67
Shoprite	-0.17	-2.45	0.72	5.75	-0.67	-4.46
SAB	-0.06	-1.40	0.92	11.29	-0.07	-0.70
Sappi	0.01	0.18	1.15	8.05	0.08	0.49
Sasol	0.09	1.29	1.01	8.36	0.41	2.79
Standard B	-0.27	-5.43	1.16	12.79	-0.36	-3.26

Figure 7 shows the efficient frontier produced from the Sharpe multiple index model. For comparison purposes, the efficient frontier from the Sharpe single index and Markowitz models have been plotted on the same graph. It is immediately noticed that the Sharpe multiple index model has shifted to the left from the Single index model. The left shift means that there has been a reduction in the model estimate of portfolio risk in moving from the Sharpe single index to the Sharpe multiple index model. There are two possible reasons for this reduction in the estimate of risk. Firstly, the beta estimates changed when multiple indices were used and secondly, the introduction of

Table 7: Composition of the Multiple Index Optimal Portfolio

Share	Markowitz	Sharpe	Tros-Hos
Firststrand	6.18%	0.97%	6.53%
Loumin (Jsc)	42.69%	42.78%	42.50%
Pick n Pay Holdings	0.53%	4.43%	0.62%
Shoprite	29.27%	26.13%	28.93%
Sasol	21.33%	25.69%	21.42%

more explanatory variables resulted in the reduction in the residual variance. Note however that despite the reduction in risk, the Sharpe multiple index model still overestimates the risk when compared to Markowitz.

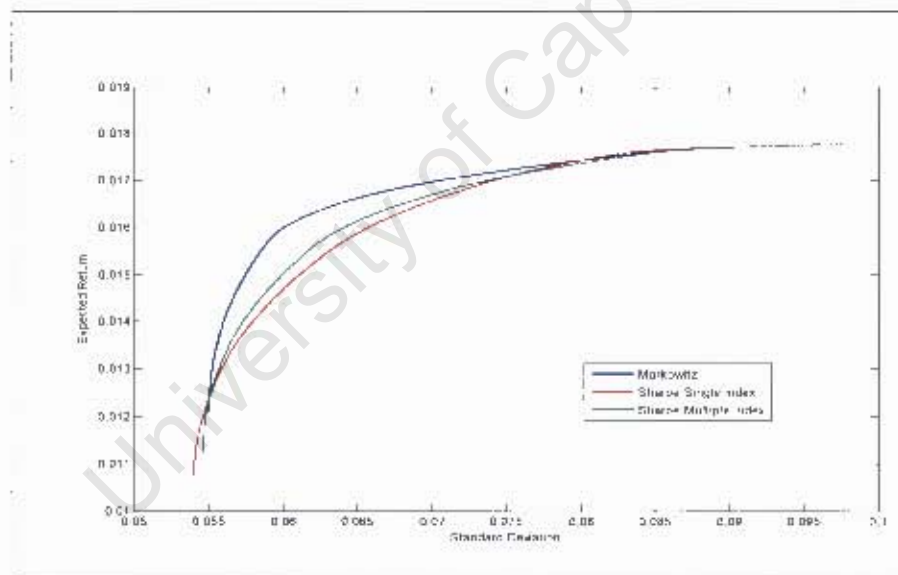


Figure 7: Sharpe Multiple Index Models

Figure 8 shows that the efficient frontier from the Troskie-Hossain multiple index model has shifted significantly to the left from the Sharpe multiple

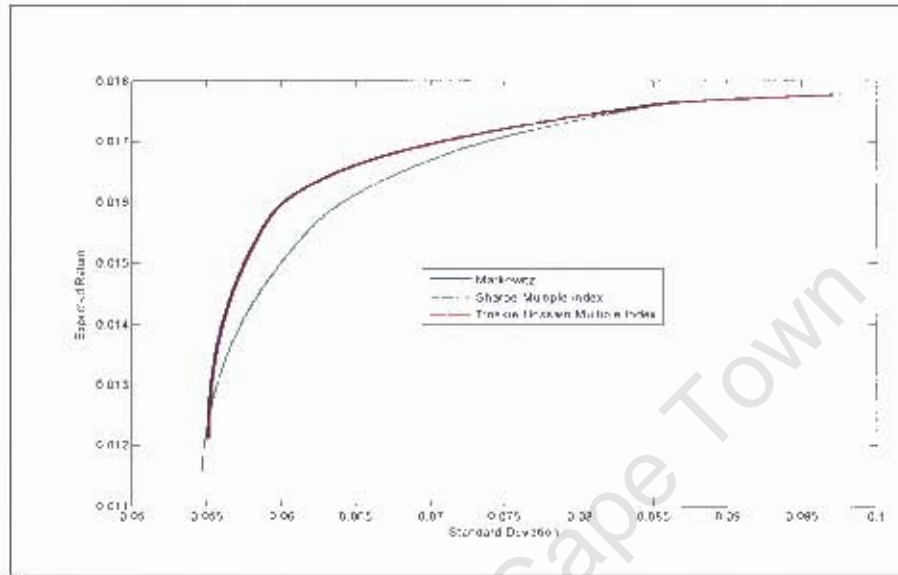


Figure 8: Troskie-Hossain Improved Multiple Index Model

Table 8: Risk-Return properties of the optimal portfolio

Method	Expected Return	Standard Deviation
Markowitz	19.8%	21.89%
Sharpe Multiple Index	19.8%	23.62%
Tros-Hos Multiple Index	19.68%	21.96%

index model. Since the only difference between these models is in the estimation of risk, it means that there has been a reduction of risk in the Troskie-Hossain model. The off-diagonal elements of the residual covariance matrix must have had a big impact which Sharpe does not take into account. Notice from Table 7 that the composition of the optimal risky portfolio from the Troskie-Hossain model is very similar to the Markowitz model.

Closer inspection of the residual correlation matrix confirms the presence of highly significant correlations. In fact, using the Bonferroni multiple t-test of significant correlation, it was found that the correlations were significantly

Figure 9: Multiple Index Residual Correlation Matrix

1.00	-0.27	-0.04	0.11	-0.01	0.02	0.39	-0.07	0.10	0.16	0.04	0.06	-0.18	-0.23	0.45
-0.27	1.00	-0.09	-0.26	0.04	-0.01	-0.36	0.20	-0.19	-0.12	-0.15	-0.11	0.24	0.26	-0.46
-0.04	-0.09	1.00	0.36	0.29	-0.16	0.03	-0.11	0.11	0.37	0.29	0.11	0.11	-0.03	-0.15
0.11	-0.26	0.36	1.00	0.20	0.06	0.36	-0.19	0.16	0.30	0.33	0.08	-0.07	-0.26	0.07
-0.01	0.04	0.29	0.20	1.00	-0.01	0.03	-0.10	0.03	0.25	0.07	0.18	0.11	0.05	0.01
0.02	-0.01	-0.16	0.06	-0.01	1.00	0.06	0.07	0.06	-0.25	0.14	-0.06	0.00	0.01	0.02
0.39	-0.36	0.03	0.36	0.03	0.06	1.00	-0.35	-0.05	0.25	0.23	0.10	-0.18	-0.30	0.39
-0.07	0.20	-0.11	-0.19	-0.10	0.07	-0.35	1.00	-0.16	-0.22	-0.26	-0.10	0.07	-0.02	-0.15
0.10	-0.19	0.11	0.16	0.03	0.06	-0.05	-0.16	1.00	0.09	0.11	-0.10	-0.11	-0.15	-0.06
0.16	-0.12	0.37	0.30	0.25	-0.25	0.25	-0.22	0.09	1.00	0.31	0.25	0.12	-0.14	0.23
0.04	-0.15	0.29	0.33	0.07	0.14	0.23	-0.26	0.11	0.31	1.00	0.06	-0.06	-0.11	-0.03
0.06	-0.11	0.11	0.08	0.18	-0.06	0.10	-0.10	-0.10	0.25	0.06	1.00	-0.08	-0.15	0.16
-0.18	0.24	0.11	-0.07	0.11	0.00	-0.18	0.07	-0.11	0.12	-0.06	-0.08	1.00	0.32	-0.08
-0.23	0.26	-0.03	-0.26	0.05	0.01	-0.30	-0.02	-0.15	-0.14	-0.11	-0.15	0.32	1.00	-0.08
0.45	-0.46	-0.15	0.07	0.01	0.02	0.39	-0.15	-0.06	0.23	-0.03	0.16	-0.08	-0.08	1.00

different from zero. In particular, most of these significant correlations were in fact negative correlations. This explains why the Sharpe multiple index model overestimates risk because these negative correlations, had they been taken into account, would have actually resulted in a reduction of risk.

3.9.5 Conclusions

The investigation using the Sharpe multiple index models revealed the following insights:

- The Sharpe multiple index model resulted in a lower residual variance than the Sharpe single index model and this contributed to the left shifting of the resulting efficient frontiers.
- The Sharpe multiple index model produced an efficient frontier that was quite different to Markowitz, the difference being that the Sharpe model had larger estimates of portfolio risk in almost all the regions of the risk-return space.
- The Troskie-Hossain multiple index model produced an efficient frontier that was quite different to the corresponding Sharpe model but which was very similar, (in fact almost identical), to the Markowitz model. This is because the covariance matrix of returns estimated using the Troskie Hossain model can be shown to be very similar to the Markowitz mle covariance estimate.
- The multiple index models showed, through the Troskie-Hossain model, that there were significant correlations between the residuals from different shares, multiple correlation tests revealed that the majority of the significant correlations were negative. These negative correlations result in the reduction of portfolio risk, causing the left shift of the efficient frontier.

3.10 Time Series Errors and Heteroskedasticity-Dynamic

Models

When the Sharpe index models were discussed in previous sections we stated the following assumptions:

$$E(e_{it}^2) = \sigma_{ei}^2 \quad (3.62)$$

$$E(e_{it}e_{is}) = 0, \quad t \neq s = 1, \dots, N, \quad (3.63)$$

$$E(e_{it}I_t) = 0, \quad t = 1, \dots, N, \quad (3.64)$$

$$E(e_{it}e_{jt}) = 0, \quad t = 1, \dots, N, \quad (3.65)$$

$$i, j = 1, \dots, p, \quad i \neq j.$$

The reason these assumptions were used is because they give elegant results for the least squares parameter estimates in terms of simple matrix algebra. Moreover, if the above assumptions are satisfied then we can use the Gauss-Markoff theory which we use introduce below.

Theorem 3.1. Gauss-Markoff

The maximum likelihood estimates of the linear model are the best linear unbiased estimates. (BLUE)

This is a very powerful result, it tells us that the estimates from the least squares method will be the best because they have minimum variance amongst all other linear unbiased estimators. For us to use the Gauss-Markoff theorem, we need to ensure that the assumptions of the regression model are satisfied. In particular, we need to ensure that the assumptions represented in Equation (3.62), constant residual variance and Equation (3.63), no autocorrelation are satisfied. If these assumptions are violated then we can no longer rely on the ordinary least squares estimates to be BLUE. Fortunately, if our regression model violates these assumptions we can modify the regression to take these factors into account. If it turns out that the residuals are

autocorrelated, then we use autoregressive (*AR*) models for the residuals and if the residual variance is non-constant (heteroskedasticity), we can use Generalized Autoregressive Conditional Heteroskedasticity (*GARCH*) models.

3.10.1 Autoregressive Models

When the regression model violates the assumption of no autocorrelation in the residuals, we can use the following model for residuals:

$$R_t = \alpha + \beta I_t + e_t \quad (3.66)$$

$$e_t = \theta_{t-1}e_{t-1} + \theta_{t-2}e_{t-2} + \dots + \theta_p e_{t-p} + v_t \quad (3.67)$$

$$E(v_t) = 0 \quad (3.68)$$

$$E(v_t^2) = \sigma_v^2 \quad (3.69)$$

This is referred to as an autoregressive model of order p , written $AR(p)$, with Equation (3.67) adjusting for the serial correlation in the residuals that could lead to unreliable estimates.

3.10.2 GARCH(p, q) Models

If the residual variance is not constant, then we need to model the residual variance using the $GARCH(p, q)$ model proposed by Engle (1982) defined as

$$\sigma^2 = \omega + \sum_{i=1}^q \theta_i e_{t-i}^2 + \sum_{i=1}^p \lambda_i \sigma_{t-i}^2. \quad (3.70)$$

Equation (3.70) is a higher order $GARCH$ model but for most practical purposes, the $GARCH(1, 1)$ defined below is used to model the residual

variance.

$$R_t = \alpha + \beta I_t + e_t \quad (3.71)$$

$$e_t = \sigma_t \varepsilon_t \quad (3.72)$$

$$\sigma_t^2 = \omega + \theta e_{t-1}^2 + \lambda \sigma_{t-1}^2. \quad (3.73)$$

$$\varepsilon_t \sim N(0, 1)$$

The GARCH(1,1) formulation thus models the conditional variance as time varying. This type of model works very well for stock market data which has shown to exhibit correlation in volatility which Cont (2001) refers to as volatility clustering. Engle (1982) and Bollerslev, Chou and Kroner (1992) have shown how this model can be used to model stock volatility and interest rates. The advantage of using GARCH(1,1) models in finance is that if the volatility is not constant, then it allows us to get the most recent estimate of the residual variance. Note that higher order models GARCH(p, q) can also be considered, but for the purpose of this study we will restrict the investigation to GARCH(1,1). The main disadvantage of using higher order GARCH(p, q) models is the considerable loss in the degrees of freedom.

3.11 Empirical Investigation of the AR and GARCH Index Models

3.11.1 Objectives of the Investigation

We would like to investigate the effect that serial correlation in the residuals and heteroskedasticity in the residual variance has on the efficient frontier. We will focus on the Sharpe single index model which we estimate using the least squares method. We will start by estimating the regression using the AR(p). What will be of particular interest is the effect of the combination of both the AR(p) and the GARCH(1, 1) model in the estimation because the aim is to get the model as close as possible to satisfying the Gauss-

Markoff theorem. This is very important because if the model satisfies the Gauss-Markoff theorem, then the regression will yield the best linear unbiased estimates.

3.11.2 Data used in the Investigation

The data used will be the same as the data used in the Sharpe single index empirical investigation.

3.11.3 Method of Investigation

The first step in the process involves using the ordinary least squares regression results to check if the assumptions have been violated. For the $AR(p)$ models, we need to test for significant autocorrelation in the residuals and for the GARCH(1, 1) model, we test if there is significant autocorrelation in the *squares* of the residuals when we have removed the autocorrelation in the residuals. Eviews gives us a very convenient way of doing both these tests using the Q-statistic. If significant $AR(p)$ terms are found then we add the relevant $AR(p)$ term in the regression and the regression is performed. The investigation is performed on following models:

1. The Sharpe single index $AR(p)$ models.
2. The Sharpe single index $AR(p)$ -GARCH(1, 1) model.

Since serial autocorrelation and heteroskedasticity are now taken into account, the resulting regression estimates are different to the least squares estimates. We use the new estimates to compute the mean vector and covariance matrix to use in the efficient frontier.

3.11.4 Results

Table 9 shows that many of the shares in the portfolio had significant autocorrelation which would violate the regression assumptions. The effect this

Table 9: Autocorrelation

Share	AR(p)	t-Statistic
Absa	-	-
Anglos	AR(2)	-1.95
Barlword	AR(1)	-1.90
Bidvest	-	-
Group 5	-	-
Illovo	AR(2)	-1.96
Firststrand	-	-
Lonmin	AR(3)	2.73
Mr price	AR(3)	2.56
Pick n pay	AR(1)	-3.75
Shoprite	-	-
SAB	AR(1)	-2.27
Sappi	-	-
Sasol	-	-
Standard B	-	-

autocorrelation has on the efficient frontier is shown in Figure 10.

The efficient frontier shows very little difference between the efficient frontier estimated using AR terms and the ordinary least squares. The only real difference is in the extreme regions of the risk-return space where the Sharpe model is more efficient than the AR model. The composition of the resulting optimal risky portfolios of both methods are identical. This result means that for portfolio analysis, the presence of autocorrelation alone may not have a serious impact on the efficient frontier.

Table 10 show the shares that had significant $GARCH(1, 1)$ terms. These terms are as a result of significant correlation in the squared residuals which violates the assumption that the errors should be independent. It would be very interesting to see how this affects the portfolio analysis so we will fit the relevant GARCH terms and perform the portfolio analysis. This is exactly what is reported by Hossain, Troskie and Guo, (2005).

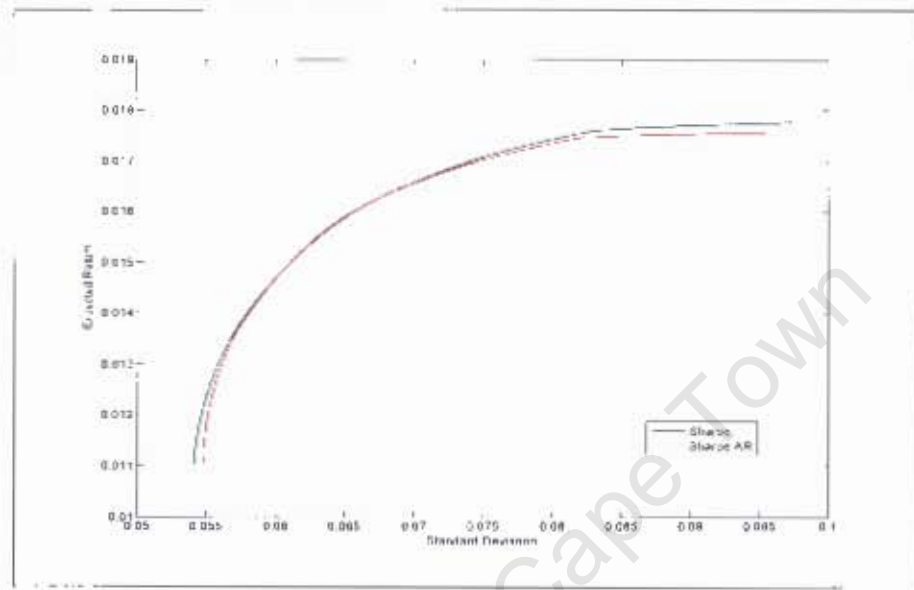


Figure 10: Sharpe AR Model Efficient Frontier

Table 10: Heteroskedasticity

Share	GARCH
Absa	GARCH(1,1)
Anglos	GARCH(1,1)
Barlword	GARCH(1,1)
Bidvest	-
Group 5	-
Illovo	-
Firstrand	-
Lonmin	GARCH(1,1)
Mr price	GARCH(1,1)
Pick n pay	GARCH(1,1)
Shoprite	GARCH(1,1)
SAB	GARCH(1,1)
Sappi	-
Sasol	-
Standard B	GARCH(1,1)

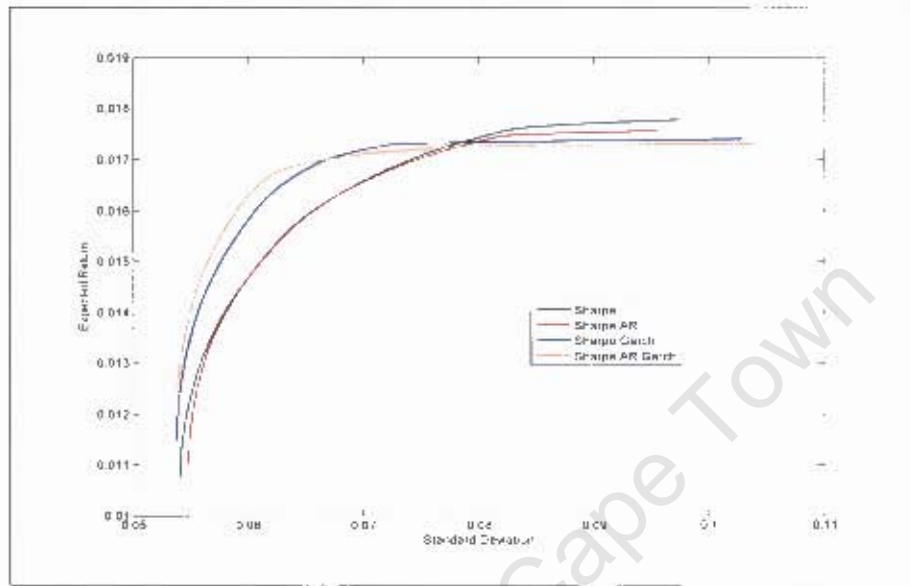


Figure 11: AR-GARCH Efficient Frontier

In Figure 11, the efficient frontier from the AR-GARCH model is shown. This model combines both the AR and the GARCH(1,1) terms in the estimation. It is very surprising to note that although the AR model resulted in a right shift of the efficient frontier, when it is combined with the GARCH model, the efficient frontier shifts further to the left of the GARCH model. In other words, the combination of AR-GARCH model is more efficient than each of the separate models. This can be justified by the fact that the model has taken both autocorrelation and heteroskedasticity into account meaning that it is much closer to satisfying the Gauss-Markoff theorem.

3.11.5 Conclusions

The investigation using the $AR(p)$ and GARCH(1,1) Sharpe single index models revealed the following insights:

- Autocorrelation in the residuals violates the regression assumptions and share returns exhibit a considerable amount of autocorrelation when estimated using the Sharpe single index model.
- The presence of autocorrelation made a very small difference to the efficient frontiers.
- Share returns can show autocorrelation in the squares of the residuals, meaning that even if the residuals are uncorrelated, they are not independent since the *squares* can be autocorrelated.
- The combination of both $AR(p)$ and $GARCH(1, 1)$ terms in the regression gives the "best" results because by correcting for both the autocorrelation and heteroskedasticity, the residuals will be much closer to satisfying the Gauss-Markov theorem.

4 State Space Models with the Kalman Filter

4.1 State Space Analysis

Suppose we have a set of observations $\mathbf{Y} = \{y_1, y_2, y_3, \dots, y_n\}$ that are associated with a series of *unobserved* $\mathbf{B} = \{\beta_1, \beta_2, \beta_3, \dots, \beta_n\}$. State space techniques give us a way of determining the properties of the unobserved variable whenever we observe \mathbf{Y} . We would like to use state space modelling techniques to model financial variables that are not observed in the market. For example, on the Johannesburg Stock Exchange, we can observe the stock prices and hence the returns. If we decide to use the Sharpe single index model as defined in the previous chapter, we have

$$R_{it} = \alpha_i + \beta_i I_t + e_{it}. \quad (4.1)$$

We can observe R_{it} and I_t but the β_i is not a market observable quantity so we may decide to model the β_i as a state variable. This is precisely what we intend to do later in this chapter because one of the main aims of this thesis is to provide dynamic estimates of risk and return using these unobserved quantities. For now, we will start this chapter by developing the mathematical theory behind state space modelling using the Kalman Filter.

4.2 The Kalman Filtering Algorithm

The Kalman Filter is a recursive estimation algorithm for state variable estimation that was developed by Rudolph E. Kalman, (Kalman, 1960) for discrete data linear filtering problems. The method has been used primarily in engineering applications, in problems involving tracking systems and motion prediction. In financial applications, it has been used to model unobserved macroeconomic variables such as inflation by Burmeister et al., (1986). In

presenting the Kalman filter, we will use the general multivariate case used by Shumway and Stoffer (2000 p. 313).

Let \mathbf{Y}_t be the vector of observations available at time t . The linear model that Shumway and Stoffer (2000 p. 313) consider has the following form,

$$\mathbf{Y}_t = \mathbf{A}_t \mathbf{x}_t + \mathbf{v}_t, \quad \mathbf{v}_t \sim N(\mathbf{0}, \mathbf{R}) \quad (4.2)$$

$$\mathbf{x}_t = \Phi \mathbf{x}_{t-1} + \mathbf{w}_t, \quad \mathbf{w}_t \sim N(\mathbf{0}, \mathbf{Q}) \quad (4.3)$$

Equation (4.2) defines the signal or measurement equation and Equation (4.3) is the state equation. The state variable \mathbf{x}_t is unobserved. We would like to investigate the properties of \mathbf{x}_t everytime a new measurement \mathbf{Y}_t becomes available. From equation (4.3), we have

$$E(\mathbf{x}_t | \mathbf{Y}_{t-1}) = \Phi \mathbf{x}_{t-1} \quad (4.4)$$

This is the prior estimate of \mathbf{x}_t before \mathbf{Y}_t is observed. If we let

$$\begin{aligned} \mathbf{P}_t^{t-1} &= \text{Var}(\mathbf{x}_t | \mathbf{Y}_{t-1}) \\ &= \Phi \mathbf{P}_{t-1}^{t-1} \Phi' + \mathbf{Q} \end{aligned} \quad (4.5)$$

Equation (4.5) represents the prior variance of the state variable \mathbf{x}_t before \mathbf{Y}_t is observed. If at time $t-1$ we make a prediction of \mathbf{Y}_{t-1} , then the error in the prediction \mathbf{e}_t is given by,

$$\mathbf{e}_t = \mathbf{Y}_t - \mathbf{A}_t \Phi \mathbf{x}_{t-1}$$

The Kalman gain \mathbf{K}_t is shown by Shumway (1988 p.177) as the quantity,

$$\mathbf{K}_t = \mathbf{P}_t^{t-1} \mathbf{A}_t' (\mathbf{A}_t \mathbf{P}_t^{t-1} \mathbf{A}_t' + \mathbf{R})^{-1} \quad (4.6)$$

The posterior mean and variance of the state variable \mathbf{x}_t after \mathbf{Y}_t is observed is given by Shumway (1988) as

$$E(\mathbf{x}_t|\mathbf{Y}_t) = \Phi\mathbf{x}_{t-1} + \mathbf{K}_t\mathbf{e}_t \quad (4.7)$$

$$Var(\mathbf{x}_t|\mathbf{Y}_t) = \mathbf{P}_t^{t-1} - \mathbf{K}_t\mathbf{A}_t\mathbf{P}_t^{t-1} \quad (4.8)$$

To summarize, the Kalman filtering process thus has the following stages.

1. Assume an initial distribution of the state variable, in our case, we let $(\mathbf{x}_{t-1}|\mathbf{Y}_{t-1}) \sim N(\Phi\mathbf{x}_{t-2}, \mathbf{P}_{t-1}^{t-1})$.
2. Make a prediction at time $t - 1$ about the time t value of the state variable; in our case the best estimate we can make is $E(\mathbf{x}_t|\mathbf{Y}_{t-1}) = \Phi\mathbf{x}_{t-1}$
3. Using the signal equation $\mathbf{Y}_t = \mathbf{A}_t\mathbf{x}_t + \mathbf{v}_t$, make a prediction of $\hat{\mathbf{Y}}_t$ using $\hat{\mathbf{Y}}_t = \mathbf{A}_t\Phi\mathbf{x}_{t-1}$.
4. Now at time t look at the actual observed value of \mathbf{Y}_t and compared with $\hat{\mathbf{Y}}_t$ then calculate the error $\mathbf{e}_t = \mathbf{Y}_t - \hat{\mathbf{Y}}_t$
5. Now update (filter) the posterior distribution as using the error from the previous step, ie update it to get $E(\mathbf{x}_t|\mathbf{Y}_t) = \Phi\mathbf{x}_{t-1} + \mathbf{K}_t\mathbf{e}_t$ and $Var(\mathbf{x}_t|\mathbf{Y}_t) = \mathbf{P}_t^{t-1} - \mathbf{K}_t\mathbf{A}_t\mathbf{P}_t^{t-1}$ which we can write as $(\mathbf{x}_t|\mathbf{Y}_t) \sim N(\Phi\mathbf{x}_{t-1}, \mathbf{P}_t^t)$
6. Now go back to step 1 and repeat.

4.3 Rationale behind the Kalman Filter

Before we use the Kalman Filter we need to investigate several issues. The first issue is that of convergence. In other words, how can we be sure that the Kalman Filter actually filters the noise from the system such that after a large

number of steps, the estimate of the state variable is close to the true value? This is important, because if the Kalman Filter never converges, then it will give unreliable results. The derivation of the Kalman Gain that we used was from a purely statistical standpoint. However in the original derivation by Kalman, (1960), the Kalman gain is chosen such that it minimizes the variance of the posterior state variable. In other words, each time we filter, a correction is made to the prior state variable that would minimize the posterior state variance.

Another interesting feature of the Kalman filter is that it gives more weighting to the most recent observation. This is a useful feature because it allows the filter to use more recent estimates of the parameters. For example, ordinary least squares gives equal weighting to all data points and this may not give the best estimate of the current state. In later sections, we will be using the Kalman Filter in financial applications and this feature of weighting of observations will prove to be very useful in dynamic estimation of risk.

The usefulness of the Kalman Filter is summarized by Maybeck, (1979 pg. 4) as follows: "... it processes all available measurements regardless of their precision, to estimate the current value of the variables of interest, with use of (1) knowledge of the system and measurement device dynamics, (2) the statistical description of the system noises, measurement errors, and uncertainty in the dynamic models, and (3) any available information about the initial conditions of the variables of interest."

4.4 Estimation of the Parameters

In the derivation of the Kalman Filter above, we assumed that the parameters are known, but this is not the case, in practice we actually need to estimate these parameters from the data. Conveniently, since we have assumed normality, we can use the maximum likelihood estimates. Assuming

that the signal (observation) and state equations are given by,

$$\mathbf{Y}_t = \mathbf{A}_t \mathbf{x}_t + \mathbf{v}_t, \quad \mathbf{v}_t \sim N(\mathbf{0}, \mathbf{R}) \quad (4.9)$$

$$\mathbf{x}_t = \Phi \mathbf{x}_{t-1} + \mathbf{w}_t, \quad \mathbf{w}_t \sim N(\mathbf{0}, \mathbf{Q}) \quad (4.10)$$

Shumway and Stoffer (2000 p. 321) show that conditional on \mathbf{Y}_{t-1} , the one step ahead error is distributed,

$$\mathbf{e}_t | \mathbf{Y}_{t-1} = \mathbf{Y}_t - \mathbf{A}_t \Phi \mathbf{x}_{t-1} \sim N(\mathbf{0}, \mathbf{A}_t \mathbf{P}_t^{t-1} \mathbf{A}_t' + \mathbf{R}) \quad (4.11)$$

The parameters \mathbf{R} , Φ and \mathbf{Q} can then be estimated by the method of maximum likelihood.

4.5 Smoothed State Estimates

The Kalman filter can be very sensitive to outlying observations and because it gives the most weight to the most recent observation, if that observation is outlying, then the current state variable will also be contaminated. In addition, because the Kalman Filter is a discrete estimation algorithm, it assumes that the state variable only changes when a new observation is introduced. In practice, the state variable changes continuously and so it would be very convenient if we could adjust the estimates in order to smooth them out. This could help in clearly identifying any trends in the state variables, enabling better forecasting. In order to smooth out the state estimates, a backward recursion algorithm, such as the one suggested in Shumway (1988 p.177) is often used. Smoothing is very important especially when dealing with macro-economic variables in finance. For example, if we model the beta estimate of the Sharpe single index model, we expect that if this beta changes, then it does so gradually instead of changing at discrete time points. The reason we will sometimes use smoothed estimates is because smoothing indicates

gradual changes in the state variables which is what we expect to happen in practice.

4.6 Application of The Kalman Filter to the Single

Index Models

The derivation of the Kalman Filter gave important insights into the properties, assumptions and estimation techniques involved in state space modelling. The recursive nature of the Kalman filter allows the development of dynamic time-varying models. The Sharpe single index model assumes that the regression coefficients are constant. However in practice, there is evidence that the betas change over time. Bowie, (1994) has explored various methods of accounting for changes in the betas using dynamic models. We will focus on using the Sharpe single index model as a basis for developing dynamic models through the Kalman filter. Recall that the Sharpe single index model uses the following relationship between the returns R_{it} and the market proxy index I_t . Note that there is no time index on the beta since it is assumed that beta is constant with respect to time

$$R_{it} = \alpha_i + \beta_i I_t + e_{it} \quad (4.12)$$

Using the Sharpe single index model but modeling β_i as a time-varying unobservable state variable, the dynamic single index model becomes,

$$R_{it} = \alpha_i + \beta_{it} I_t + e_{it} \quad (4.13)$$

$$\beta_{it} = \beta_{it-1} + w_{it} \quad (4.14)$$

$$i = 1 \dots p \text{ and } t = 1 \dots N.$$

The above specification is similar to the state space specification in Equation (4.9) with

$$\begin{aligned}
 \mathbf{Y}_t &= \mathbf{A}_t \mathbf{x}_t + \mathbf{v}_t \\
 R_{it} &= (\alpha_i, I_t) \begin{pmatrix} 1 \\ \beta_{it} \end{pmatrix} + e_{it} \text{ and} \\
 \mathbf{x}_t &= \mathbf{\Phi} \mathbf{x}_{t-1} + \mathbf{w}_t \\
 \beta_{it} &= (1) \beta_{it-1} + w_{it}
 \end{aligned}$$

This is essentially the time varying capital asset pricing model (CAPM) with varying slope. Tsay (2005 p.510). Equation 4.13 is referred to as the signal equation while equation 4.14 is known as the state equation. If $E(I_t) = \mu_I$ and $var(I_t) = \sigma_I^2$ is the expected return and variance of the market returns then the expected return from share i , is given by

$$E_{it} = E(R_{it}) = \alpha_i + \beta_{it} \mu_I \quad (4.15)$$

The variance is given by

$$\begin{aligned}
 var(R_{it}) &= var(\alpha_i + \beta_{it} I_t + e_{it}) \\
 &= \beta_{it}^2 \sigma_I^2 + \sigma_{ei}^2 \\
 &= \sigma_{it}^2
 \end{aligned} \quad (4.16)$$

Both results look similar to the Sharpe formulation but note that the beta estimates β_{it} are now time dependent. This is very useful because we now have a model that is able to give dynamic estimates of mean and variance. In other words, at each time step we can evaluate the one step ahead forecast

of the mean and covariance matrix. The covariance structure is given by

$$\begin{aligned} \text{cov}(R_{it}R_{jt}) &= \beta_{it}\beta_{jt}\sigma_I^2 & (4.17) \\ &= \sigma_{ijt} \\ &i \neq j \end{aligned}$$

Now that we have estimates for the expected returns and covariance matrix, we can go directly to the optimization procedure to get the efficient frontier. If μ_{pt} is the expected return and σ_{pt}^2 is the variance of a portfolio at time t then the procedure we apply to get the efficient frontier is as follows:

$$\text{Minimize } \sigma_{pt}^2 = \mathbf{w}^T \boldsymbol{\Sigma}_t \mathbf{w} \quad (4.18)$$

$$\text{such that } \mu_{pt} = \mathbf{w}^T \boldsymbol{\mu}_t \quad (4.19)$$

$$\text{and } 0 \leq w_i \leq 1.$$

The expected return and variance of the portfolio are defined as,

$$\begin{aligned} \mu_{pt} &= \sum_{i=1}^p w_i E_{it} \\ &= \sum_{i=1}^p w_i (\alpha_i + \beta_{it} \mu_I) \\ \sigma_{pt}^2 &= \sum_{i=1}^p \sum_{j=1}^p w_i w_j \sigma_{ijt}. \end{aligned}$$

Notice that in this case, the inputs to the portfolio optimization are time dependent so we get a series of efficient frontiers for every time point. In our investigation, we are interested in the most recent estimate of the efficient frontier so we will use the one step ahead estimate of the betas.

4.7 The Dynamic Mean Model

Instead of using the Sharpe model in which the share returns are related to the market, we propose to use the following dynamic model for the share returns.

$$R_{it} = \mu_{it} + e_{it} \quad (4.20)$$

$$\mu_{it+1} = \mu_{it} + w_{it} \quad (4.21)$$

$$\text{Var}(e_{it}) = R \quad (4.22)$$

The model is called *the Structural Equation Model* Harvey (1990 p. 19). It is also referred to as the *local level model* and it has been used by Durbin and Koopman, (2001) to model the annual flow volume of the Nile river. We will use this model for the asset returns. We can estimate the above model quite easily using the Kalman filter, treating Equation 4.20 as the signal equation. Note that unlike the Sharpe model, we do not make any assumption about the relationship between the share returns and market returns. In addition, we will assume that the residuals between different shares are correlated. In Equation (4.20), we are treating the mean return μ_{it} as the dynamic time varying state variable which follows a random walk. When we fit the above model, our aim is to get estimates for the vector of means and the covariance matrix which we can then use in portfolio optimization. To estimate the vector of expected returns at time t , we simply use the most recent estimate of μ_{it} for each share. For the covariance matrix of returns, we use the estimate of the variance of the residual in the signal equation. Using the notation from equation (4.22), we use \mathbf{R} . This gives us a complete set of inputs for the portfolio selection model. It will be very interesting to see how this model performs compared to the Markowitz model because like the Markowitz model, there is no need to specify the relationship between asset returns and the market index. This reduces the number of parameters to be estimated by the model because we no longer need to estimate betas which

means we lose less degrees of freedom. A further advantage of using the dynamic mean model is that we do not need to have estimates for the expected returns and variance of the market index which means that the sampling error associated with these estimators is not brought into the model.

4.8 Empirical Investigation of The Kalman Filter in Portfolio Theory

4.8.1 Objectives of the Investigation

The Kalman filter gives us a method of dynamic estimation of unobserved variables and we would like to apply this method of estimation to the Sharpe Index Models. The investigation will start by comparing the results from the ordinary least squares estimation of the Sharpe single index models to the dynamic estimation using the Kalman filter. What will be of particular interest is to investigate how the dynamic beta estimates change over time because this could certainly add value in dynamic portfolio construction. The proposed model for asset returns that treats the mean as the state variable (Equation 4.20) will also be investigated and compared to the Markowitz model. The investigation will conclude by evaluating the difference between the efficient frontiers constructed from dynamic betas to those constructed from ordinary least squares betas.

4.8.2 Method of Investigation

The dynamic estimation of the Sharpe index model is performed in the Eviews package using the state space method. Eviews gives a great deal of flexibility in terms of the model specification of the dynamics for the signal and state equations. We will focus on the time-varying estimation method that enables us to model the beta estimates as a random walk. We could have chosen to model the betas with autoregressive type models such as

$$\beta_{it} = c + \phi\beta_{it-1} + w_{it} \quad (4.23)$$

with c and ϕ constants.

but preliminary investigations showed that the random walk model represented by

$$\beta_{it} = \beta_{it-1} + w_{it}, \quad (4.24)$$

provides a much better fit with a lower residual variance in the signal equation. The random walk model for the betas was used as the beta dynamics for all the shares. Once the model parameters are estimated in Eviews, the beta estimates are used as inputs in the Sharpe index models, the only difference is that instead of using the constant beta estimates from the least squares estimates, we use the one step ahead forecasts of the betas from the Kalman filter. The one step ahead betas, together with the residual variance from the signal equation, are used to construct the one step ahead mean and covariance matrix. The efficient frontiers are then constructed using these inputs in Matlab7.

In addition to using the Kalman filter in the Sharpe single index model, the following model was estimated in Eviews. This is a well known structural equation model Harvey (1990 p. 19) and it treats the mean return as the dynamic time-varying state variable.

$$\begin{aligned} R_{it} &= \mu_{it} + e_{it} \\ \mu_{it+1} &= \mu_{it} + w_{it} \\ \text{Var}(e_{it}) &= R \end{aligned} \quad (4.25)$$

The inputs for the portfolio optimization from this model were then used

in Matlab7 to construct the efficient frontier.

4.8.3 Results

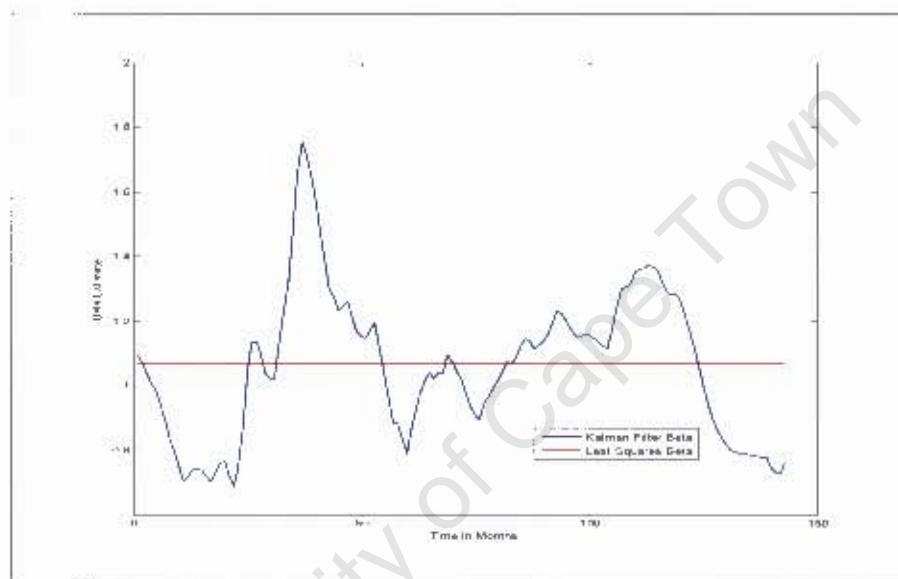


Figure 12: Kalman Filter Estimates of Beta

Figure 12 shows the time-varying beta for the share Sasol, estimated from the Kalman filter (blue curve) plotted together with the least squares estimate of beta shown as the red line. The least squares estimation assumes that the beta is constant over the estimation period hence it is shown as a constant flat line. The Kalman filter does not make that assumption and so it gives time varying estimates, clearly showing that the beta is in fact not constant. What is interesting is that the dynamic beta estimates are often much higher or much lower than the least squares estimates for extended periods of time. Using the constant least squares estimates will thus lead to grossly under or overestimation of the beta. This can have a disastrous effect in terms of managing risk because a fund manager using least squares

Table 11: Least Squares vs Kalman Filter Estimates

Share	β_{LS}	β_{Kal}	$\sigma_{ei(LS)}$	$\sigma_{ei(Kal)}$	R^2_{LS}	R^2_{Kal}
Absa	1.07	1.02	0.082	0.075	40%	50%
Anglos	1.30	1.58	0.058	0.052	66%	73%
Barlword	1.00	1.02	0.069	0.064	45%	52%
Bidvest	0.78	1.14	0.059	0.055	40%	48%
Group 5	0.72	1.73	0.128	0.122	11%	18%
Illovo	0.72	1.04	0.078	0.074	25%	33%
Firststrand	1.02	1.33	0.078	0.069	40%	51%
Lonmin	0.87	1.14	0.081	0.073	31%	44%
Mr price	0.94	0.94	0.118	0.115	20%	23%
Pick n pay	0.62	0.50	0.087	0.081	17%	27%
Shoprite	0.60	1.37	0.087	0.082	15%	25%
SAB	0.86	0.66	0.053	0.047	51%	60%
Sappi	1.15	1.49	0.092	0.082	38%	50%
Sasol	1.06	0.76	0.080	0.075	40%	48%
Standard B	0.94	1.09	0.067	0.061	43%	52%

estimates may think that he has a low beta stock when in fact the beta at that time is very high. What is most useful about dynamic estimation of the beta is that it also gives the manager an idea of the direction of the betas in terms of whether or not the beta has peaked. As an example, looking at the very last estimate of the dynamic beta in Figure 12, it is clear that the final estimate of beta is quite low suggesting the start of an upward trend. These signals are very important because what is of greatest importance to the manager is the future value of beta not the historical value.

The dynamic estimation of the Sharpe single index model using the Kalman filter gave estimates for the regression parameters shown in table 11. The surprising result is that some of the beta estimates from the Kalman filter are much higher than the least squares estimates while some are nearly identical to the least squares. The Kalman filter model also resulted in a much higher R^2 statistic than the least squares estimation for all the shares.

These results will have a significant impact on the estimates of risk relative to the market. The much higher beta estimates from the Kalman filter mean that the shares have a much higher market risk than the least squares estimate would suggest. An investor that uses the least squares estimates of beta would underestimate the sensitivity of the share returns to the market returns over that period. It is important to bear in mind that the difference between the beta estimates is because the least squares estimate assumes that the betas are constant over the period of investigation while the Kalman filter allows the beta to change with time, giving the most recent estimate. The dynamic nature of the Kalman filter estimates ensures that the most recent observation has the most weighting meaning that it responds much quicker to changes in the true betas.

All of the shares resulted in a much lower residual variance than the least squares estimates as shown in table 11 . The residual variance is a measure of the stock specific risk which cannot be explained by the market index, meaning that by lowering the residual variance, the Kalman filter is able to explain a higher proportion of the variation in the share returns than the least squares estimates. The reduction in the residual variance together with the changes in the beta estimates will have a significant impact on portfolio selection since these are the key inputs when we use the index model. While the reduction in the residual variance from the Kalman Filter will reduce the stock specific risk, the higher betas will increase the market or systematic risk, so it is not quite clear what the overall impact will be on the portfolio. To investigate this issue further, the efficient frontier using the Kalman filter is constructed and shown in figure 13.

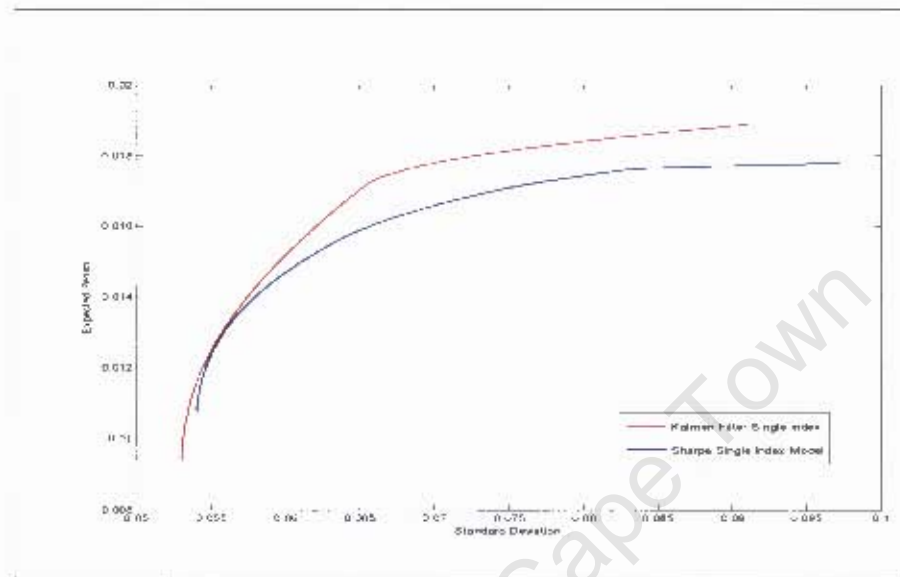


Figure 13: Kalman Filter Single Index Model

The Kalman filter is shown in figure 13 and it is clear that it is more efficient than the Sharpe least squares single index model. In the low risk region of the efficient frontier, the Sharpe least squares model is almost similar to the Kalman filter model, however as the risk gradually increases, the Kalman filter outperforms the Sharpe model. The efficient frontier has shifted up possibly due to the fact that the beta estimates are much higher than the least squares estimates which would tend to increase the expected returns. It is however important to note that the efficient frontier presented is the most recent one. The dynamic nature of the betas mean that we can construct a different efficient frontier for each time period.

Table 12: Composition of the Kalman Filter Optimal Portfolio

Share	Markowitz	Sharpe	Kalman Filter
Firststrand	6.18%	0%	16.54%
Lonmin (Jse)	42.69%	45.51%	52.70%
Pick n Pay Holdings	0.53%	2.46%	21.30%
Shoprite	29.27%	20.83%	0%
Sasol	21.33%	31.2%	0%
Standard Bank	0%	0%	9.45%

Table 13: The Optimal Risky Portfolio

Method	Expected Return	Standard Deviation
Markowitz	19.8%	21.89%
Sharpe	20.16%	25.01%
Kalman Filter	25.92%	26.78%

Table 12 shows the composition of the optimal risky portfolio estimated from the Kalman filter single index model. The composition of this portfolio is very different to both the Sharpe and the Markowitz model. Shares such as Shoprite and Sasol that were once in both models are now not part of the Kalman filter optimal risky portfolio and in addition, Standard bank is now included. These results show the impact that changes in beta estimates can have on the composition of the optimal portfolio. What is most interesting about the Kalman filter optimal portfolio is that while it has a slightly higher standard deviation than the Sharpe model, it has a much higher expected return, which gives it a better reward to risk ratio.

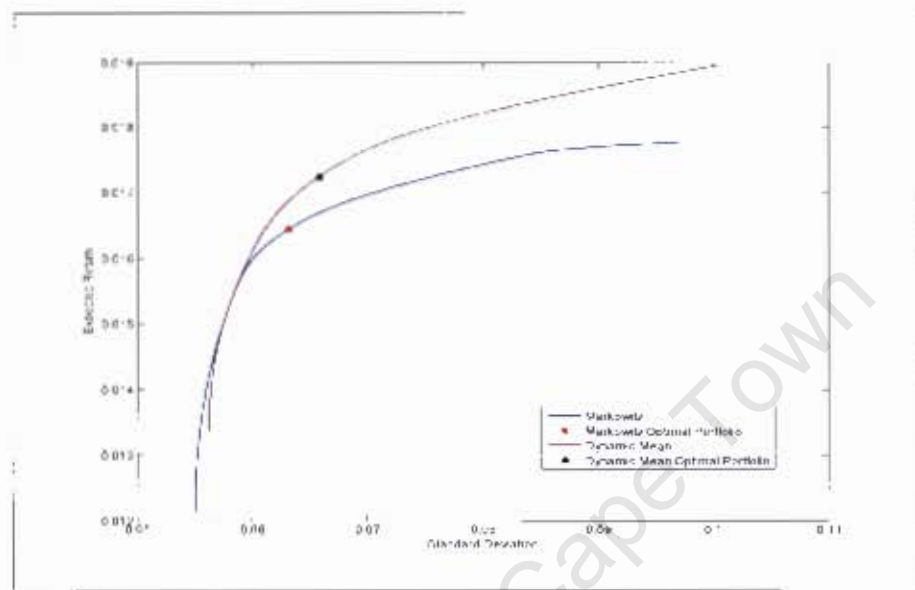


Figure 14: Dynamic Mean Efficient Frontier

Figure 14 show the efficient frontier estimated using the dynamic mean model together the Markowitz model. The efficient frontier from the dynamic mean model has shifted up compared to Markowitz. The shift is actually very similar to the shift produced by the Sharpe Kalman filter model. The optimal risky portfolio from the Dynamic Mean model has a slightly higher risk but a much higher expected return than Markowitz. In practice we will have a range of such efficient frontiers, since the estimates will depend on time. We have used the most recent efficient frontier because that is the one that is of greatest interest to the portfolio manager.

Table 14: Composition of the Dynamic Mean Optimal Portfolio

Share	Markowitz	Kalman Filter	Dynamic Mean
Group 5	0%	0%	0.49%
Firststrand	6.18%	16.54%	55.05%
Lonmin (Jse)	42.69%	52.70%	0%
Pick n Pay Holdings	0.53%	21.30%	8.19%
Shoprite	29.27%	0%	36.27
Sasol	21.33%	0%	0%
Standard Bank	0%	9.45%	0%

Table 15: The Optimal Risky Portfolio

Method	Expected Return	Standard Deviation
Markowitz	19.8%	21.89%
Kalman Filter	25.92%	26.78%
Dynamic Mean	20.64%	22.8%

There is a dramatic change in the asset allocation of the optimal risky portfolio produced by the Dynamic Mean model as shown in Table 14. What is interesting to note is that a new share, Group 5, which had never entered the optimal portfolio in previous models, is now considered. The composition of the other shares also changes drastically and Firststrand now has the greatest weighting in the optimal risky portfolio.

4.9 Conclusions

The empirical investigation using the Kalman filter in the single index model and the dynamic mean model leads to conclusions that can be summarized as follows:

- The Kalman filter, when applied to the Sharpe single index model, gives dynamic estimates of beta that show how the beta estimates change over time. The dynamic beta estimates tend to oscillate above and below the ordinary least squares estimates.
- The efficient frontier derived from the Kalman filter estimation is more efficient than the least squares Sharpe model in all regions of the efficient frontier, particularly in the higher risk regions.
- The composition of the resulting optimal risky portfolio from the Kalman filter is very different to the one from the Sharpe model. The investigation revealed that the Kalman filter produced an optimal portfolio that had a higher reward to risk ratio because it had approximately the same risk as the Sharpe model but a much higher expected return.
- Using the Dynamic Mean model to estimate inputs for portfolio selection requires less inputs than the Sharpe Index model.
- The data used in the empirical investigation showed that the Dynamic Mean model outperformed Markowitz and it had produced a similar shift to the Kalman filter Single Index Model.
- The composition of the optimal risky portfolio produced by the Dynamic Mean Model was quite different to the one produced by both the Sharpe and the Markowitz model.

Finally, the most crucial observation is that the differences in the least squares estimation and the dynamic estimation from the Kalman filter model result from the fact that in reality, the betas are not constant and so any tool that can be used to more closely track them over time may lead to better portfolio management. Similarly, we can justify using the Dynamic Mean model instead of the Markowitz model because in reality we know that asset returns are not constant over time so a model that takes this fact into account may be more appropriate.

University of Cape Town

5 Robust Portfolio Selection Techniques

5.1 Robustness and its Importance in Estimation

In previous chapters we have been using sample data to estimate statistical quantities that are of interest to portfolio theory. We used the sample estimates based on the maximum likelihood estimators from the normal distribution, and because the sample size was sufficiently large, we assumed that the estimators were good approximations of the true underlying quantities. This would be true if all our observations belonged to the same distribution but it would be wrong if the data was contaminated. In a regression setting for example, the estimates derived from data that do not satisfy the distribution assumptions due to contamination will be very unreliable. There are a variety of reasons why the data may be contaminated, it could be due to measurement errors or the presence of outliers or influential observations in the sample. Depending on the complexity of the problem, identification of these outliers may be easy but how to deal with them once identified is not as straightforward.

We would like to use models that are robust in the sense that the parameter estimates are not greatly influenced by these influential observations. Much of the pioneering work in robust estimation is due to work of Huber, (1964) and Hampel, (1968). As Huber, (1981 pg.1) puts it, "robustness signifies insensitivity to small deviations from the assumptions". This is precisely what we would like to achieve because we will use models that depend on certain assumptions, and we would like estimation methods that can resist contamination in the data. If we can use robust estimation methods, then we will have greater confidence that our sample estimates will be in the neighbourhood of the true population parameters. We will use robust methods to estimate portfolio risk and return in order to construct robust efficient

frontiers and will show that classical methods of estimating the mean and covariance matrix are very sensitive to outliers.

5.2 Multivariate Robust Estimation Theory

5.2.1 The Influence Function

The starting point for developing the theory behind robust estimation is through the notion of the so called *Influence Functions* that were proposed by Hampel, (1968). We are interested in identifying the properties of an estimator to determine how it behaves when the underlying sample from which it is derived is contaminated. Suppose that H and F_0 are distributions and consider the following set of distributions,

$$\Xi = \{F \mid F = (1 - \varepsilon)F_0 + \varepsilon H\} \quad (5.1)$$

F is thus a mixture distribution of H and F_0 and if ε is sufficiently small, then the above set is called the *contamination neighbourhood* around the model distribution F_0 with H being the contamination distribution. Huber, (1981) refers to the above formulation as the *gross error model*. Equation (5.1) tells us that each element generates realizations from F_0 with probability $(1 - \varepsilon)$ and from H with probability ε . The distribution of interest is F_0 but because of contamination, the sample may belong to some other distribution F which is an element of the above set. If T is an estimator based on the sample, then although we want to estimate $T(F_0)$, the contamination in the data means that T is now a estimator based on $T(F)$. If for example T is the maximum likelihood estimate from the distribution F_0 , then our sample estimate $T(F)$ may not give us the true estimates of the population parameters from F_0 . So we clearly need some way of determining how much of an effect this has on our estimation. One way of doing this is to use the *maximum bias* defined as,

$$b = \sup_{F \in \Xi} |T(F) - T(F_0)| \quad (5.2)$$

So the maximum bias tells us the worst case error that we can make by using an estimator based on $T(F)$ when we *should* be using $T(F_0)$. We clearly want this quantity to be as small as possible or at the very least we want it to be bounded. Similarly, the maximum variance is defined as

$$v = \sup_{F \in \Xi} \text{var}[T(F)] \quad (5.3)$$

Unfortunately as Huber, (1981) points out, equations (5.2) and (5.3) are not quite powerful enough to tell us the effect that contamination has on the data. Equation (5.1) defines a very large set indeed, we can make it more precise by limiting the set of distributions H to a special type, namely, when $H = \Delta_x$. In this case, Δ_x is a distribution that gives probability 1 to a randomly chosen point x . This gives a new neighbourhood defined by

$$\Xi_x = \{F \mid F = (1 - \varepsilon)F_0 + \varepsilon\Delta_x\} \quad (5.4)$$

The set Ξ_x now generates observations from F_0 with probability $(1 - \varepsilon)$ and observations equal to x with probability ε . This is a fundamental set in robust statistics and it is the basis of the so called *influence function* proposed by Hampel, (1968). The Influence function is defined for large samples as

$$\mathbf{IF}(x, F_0, T) = \lim_{\varepsilon \rightarrow 0} \frac{T((1 - \varepsilon)F_0 + \varepsilon\Delta_x) - T(F_0)}{\varepsilon} \quad (5.5)$$

If the derivative exists, then the influence function can be written as

$$\mathbf{IF}(x, F_0, T) = \left[\frac{\partial}{\partial \varepsilon} T(F_\varepsilon) \right]_{\varepsilon=0} \quad (5.6)$$

The influence function tells us two important properties of an estimator. It tells us not only how an infinitesimal amount of contamination affects

the estimator T but also how much an outlying observation x affects the estimator in the neighborhood Ξ_x . Indeed, as pointed out by Huber, (1981), the influence function can be considered as a function of x . We want our estimator to be robust enough to resist the outlying observation x , so it would be favourable if the estimator T has a small or a bounded influence function. The maximum value of the influence function is

$$\gamma = \sup_x |\mathbf{IF}(x, F_0, T)| \quad (5.7)$$

This tells us how outlying observations influence the estimator so we clearly want the above quantity γ to be as small as possible to ensure that our estimator is not heavily influenced (asymptotically) by outlying observations. It turns out as shown by Hampel et al. (1986), that the influence function is all we need in order to evaluate the asymptotic bias of the estimator. It turns out that an estimator that has a bounded influence function also has a bias that is asymptotically bounded. This is a very powerful result because it implies that an estimator with a bounded influence function is robust in the neighbourhood defined by Equation (5.4). Our goal is to use estimators that are resistant to outlying observations and so in order to achieve this goal, we must use estimators that have a bounded influence function.

In Chapter 2 when we discussed the Markowitz model, we used the maximum likelihood estimators $\hat{\boldsymbol{\mu}}$ ($p \times 1$) and $\hat{\boldsymbol{\Sigma}}$ ($p \times p$) assuming normality. These estimators are defined as

$$\begin{aligned} \hat{\boldsymbol{\mu}} &= \frac{1}{N} \sum_{i=1}^N \mathbf{y}_i \\ \hat{\boldsymbol{\Sigma}} &= \frac{1}{N} \sum_{i=1}^N (\mathbf{y}_i - \hat{\boldsymbol{\mu}})(\mathbf{y}_i - \hat{\boldsymbol{\mu}})' \end{aligned} \quad (5.8)$$

We blindly used the above estimators without worrying about the possibility of outliers contaminating the data. If we assume that the contamination can be modelled by Equation (5.4) then assuming that $F_0 \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, Hampel, (1986) showed quite easily that the influence functions of the maximum likelihood estimates $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}$ are given by

$$\begin{aligned}\mathbf{IF}(\mathbf{x}, F_0, \hat{\boldsymbol{\mu}}) &= -\boldsymbol{\mu} + \mathbf{x} \\ \mathbf{IF}(\mathbf{x}, F_0, \hat{\boldsymbol{\Sigma}}) &= -\boldsymbol{\Sigma} + (\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})'\end{aligned}\quad (5.9)$$

This means that the maximum likelihood estimates $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}$ both have unbounded influence functions (since \mathbf{x} can be arbitrarily large). Depending on the value of the contamination \mathbf{x} , the influence that outlying observations have on the data can be very serious which may lead to unrealistic portfolios. In fact, Victoria-Feser, (2002) showed precisely how the optimal portfolio is influenced by contamination on the data. The conclusion Victoria-Feser, (2002) makes is that contamination in both the mean and covariance leads to irrelevant portfolios since even a small amount of contamination can be given a high weight in the Markowitz optimizer.

5.2.2 The Breakdown Point of an Estimator

The influence function formulation gives the behaviour of an estimator when there is a small amount of contamination. Although this is a very useful tool for evaluating estimators, it does not tell us how they behave when there is a large amount of contamination. We thus need another measure of robustness that can tell us how estimators behave when there is a large amount of contamination. Let \mathbf{X}_d^n be an $(n \times p)$ data matrix with d of the rows of \mathbf{X} contaminated with arbitrary large values and let T be the estimator with a metric $\|\cdot\|$. The breakdown point of a location estimator as

defined in Zuo, (2001) is given by

$$B(T, \mathbf{X}) = \min \left\{ \frac{d}{n} : \sup_{\mathbf{X}_d^n} \|T(\mathbf{X}_d^n)\| = \infty \right\} \quad (5.10)$$

Intuitively, the breakdown point tells us the maximum proportion of outliers in a sample that an estimator can resist before it breaks down. In other words, the breakdown point is the highest level of contamination that an outlier can function before it starts giving arbitrarily large values. The sample median has a breakdown point of $\frac{1}{2}$ because one would need to contaminate at least half of the data before the median starts to give arbitrarily large values. The sample mean has a break down point of $\frac{1}{N}$ since it only takes a single observation to make the mean arbitrarily large. Surely, the breakdown point cannot be more than half because if more than half of the data is contaminated, it becomes difficult to identify which part of the data is outlying. The most robust estimators are those that have a high breakdown point because they lead to results that are much less influenced by outliers. We will later introduce some of these high breakdown point estimators in order to construct robust efficient frontiers.

The definition in Equation (5.10) is for location estimates such as the mean. In robust statistics, we are also interested in the properties of scale or scatter estimates such as the covariance matrix. In portfolio theory, we need both type of estimates so we want our estimate of scale to also have a high breakdown point. Huber, (1981) gives various definitions of breakdown depending on the type of estimate. We are interested in robust estimation of a square ($p \times p$) covariance matrix $C(\mathbf{X})$. Let

$$0 \leq \lambda_p(C(\mathbf{X})) \leq \lambda_{p-1}(C(\mathbf{X})) \leq \lambda_{p-2}(C(\mathbf{X})) \leq \dots \leq \lambda_1(C(\mathbf{X}))$$

be the eigenvalues of $C(\mathbf{X})$. Then Zuo, (2001) defines the breakdown point

of a scatter matrix as

$$B(C, \mathbf{X}) = \min \left\{ \frac{d}{n} : \sup_{\mathbf{X}_d^n} \left[\frac{1}{\lambda_p(C(\mathbf{X}_d^n))}, \lambda_1(C(\mathbf{X}_d^n)) \right] = \infty \right\} \quad (5.11)$$

In this case, the breakdown point is that point when either the smallest eigenvalue $\lambda_p(C(\mathbf{X}_d^n))$ is driven to zero or when the largest one $\lambda_1(C(\mathbf{X}_d^n))$ is driven to infinity.

5.2.3 Affine Equivariance

A crucial property of robust estimates that we need is the concept of *affine equivariance*. Let $\mathbf{B} = \mathbf{1}\mathbf{b}^T$ where $\mathbf{1}$ is a $(n \times 1)$ vector of 1's with a $(p \times 1)$ vector of constants \mathbf{b} and let \mathbf{A} be a $(p \times p)$ non-singular matrix. An estimator of scale, $T(\mathbf{X})$ and an estimator of scatter, $C(\mathbf{X})$ are affine equivariant if they satisfy

$$T(\mathbf{X}\mathbf{A} + \mathbf{B}) = \mathbf{A}^T T(\mathbf{X}) + \mathbf{b} \quad (5.12)$$

$$C(\mathbf{X}\mathbf{A} + \mathbf{B}) = \mathbf{A}^T C(\mathbf{X}) \mathbf{A} \quad (5.13)$$

The class of these affine equivariant estimators is very important because equations (5.12) and (5.13) define properties that we would intuitively expect a good estimator to have. The robust estimators we use for the portfolio optimization possess this crucial property.

5.2.4 Consistency and Efficiency

Although robust estimators are designed to be resistant to contamination, they still need to satisfy other statistical properties if they are to compete with other classical estimators such as the maximum likelihood estimates.

The first property is that of *consistency*, which means that the estimator must converge in probability to the quantity it is trying to estimate. The other property we want the robust estimator to have is that of *efficiency*, meaning that we want it to have minimum variance. Many of the robust estimators we will use have minor modifications designed specifically to improve their consistency and efficiency.

University of Cape Town

5.3 The Minimum Covariance Determinant Estimator

The Minimum Covariance Determinant (MCD) proposed by Rousseeuw, (1984) is a robust estimator of location and scale for multivariate data. Its robustness is due to its high breakdown point and it has the useful property of consistency. Given that the data set is an $(n \times p)$ matrix \mathbf{X} , the MCD estimator is defined for a given h , $(\frac{n}{2} < h < n)$ as the classical covariance matrix of the subset $(h \times p)$ matrix of \mathbf{X} that has the lowest covariance determinant. A more formal definition of the MCD estimator is given below,

Definition 5.1 Let $p \leq \frac{n}{2}$, and $X = \{x_1, \dots, x_n\}$ be a set of points in R^p . Let h be a natural number, $(\frac{n}{2} < h < n)$. The Minimum Covariance Determinant problem for X and h , (MCD for short), is the problem to find an h -element set $X' = \{x_1, \dots, x_h\} \subset X$ such that the determinant of the empirical covariance matrix of X' , $(\det(\text{cov}(X')))$ is minimal over all h -element subsets.

The raw-MCD estimator of location $T(X') = \hat{\mu}_{MCD}$ is just the average of the h points of X' and the raw-MCD estimator of scale $C(X') = \hat{\Sigma}_{MCD}$ is the covariance matrix of these points multiplied by a factor to improve consistency. The raw MCD estimators of location and scale, for a given h can thus be represented as the pair

$$\begin{aligned}\hat{\mu}_{MCD} &= \frac{1}{h} \sum_{x \in X'} x \\ \hat{\Sigma}_{MCD} &= a \times \text{Cov}(X')\end{aligned}\tag{5.14}$$

with the constant a chosen to achieve finite sample consistency. The MCD is robust because by choosing the matrix of minimum covariance determinant X' , it deliberately ignores any outlying observations which would have otherwise made the determinant of the covariance matrix arbitrarily large. Apart from being a robust estimator, the benefit of using the MCD

estimator is that one can choose the number h depending on the suspected level of contamination. The number h is very important because the proportion $(1 - \frac{h}{n})$, represents the amount of contamination the outlier can resist. If the data set is highly contaminated, one chooses a low h close to $\frac{n}{2}$ making the estimator more resistant to outliers. Although the raw-MCD estimator has a high breakdown point, Rousseeuw, (1999) points out that it is not very efficient for small samples hence a weighted MCD is preferred. The weighted MCD estimator is computed from the following weights,

$$\begin{aligned} w_i &= 1 \text{ if } (x_i - \hat{\mu}_{MCD})^T \hat{\Sigma}_{MCD}^{-1} (x_i - \hat{\mu}_{MCD}) \leq \chi_{p,0.975}^2 \\ w_i &= 0 \text{ if } (x_i - \hat{\mu}_{MCD})^T \hat{\Sigma}_{MCD}^{-1} (x_i - \hat{\mu}_{MCD}) > \chi_{p,0.975}^2 \end{aligned} \quad (5.16)$$

The above weights are used to calculate the reweighted MCD estimators $\hat{\mu}_{MCD}^*$ and as $\hat{\Sigma}_{MCD}^*$

$$\hat{\mu}_{MCD}^* = \frac{\sum_{i=1}^n w_i x_i}{\sum_{i=1}^n w_i} \quad (5.17)$$

and

$$\hat{\Sigma}_{MCD}^* = \frac{\sum_{i=1}^n w_i (x_i - \hat{\mu}_{MCD}^*)^T (x_i - \hat{\mu}_{MCD}^*)}{\sum_{i=1}^n w_i - 1} \quad (5.18)$$

The reweighted MCD estimator uses the raw estimators $\hat{\mu}_{MCD}$ and $\hat{\Sigma}_{MCD}$ to find the weights w_i for each observation by looking at their robust distance. Rousseeuw, (1987) adds that the raw estimates should be used only as a way to detect the outliers using Equation (5.16) and then one should use the weights to re-calculate the MLE estimates on the clean sample using Equations (5.17) and (5.18).

5.4 Empirical Investigation of the Minimum Covariance Determinant Estimator

5.4.1 Objectives of the Investigation

We now have a tool for robust estimation of the mean and covariance of returns for a multivariate data set represented by an $(n \times p)$ matrix \mathbf{X} . In our case the \mathbf{X} matrix is the returns matrix. We propose to use the robust MCD estimators as the inputs in the Markowitz portfolio optimization problem. The primary objective will be to compare how the resulting efficient frontiers from both the raw-MCD and the reweighted-MCD estimators compare to the Markowitz formulation. We would also want to compare how the optimal risky portfolios differ when robust estimators are used. The MCD estimator allows us to identify the outlying observations through the measure of robust distance defined as

$$RD_i = \sqrt{(x_i - \hat{\mu}_{MCD})^T \hat{\Sigma}_{MCD}^{-1} (x_i - \hat{\mu}_{MCD})} \quad (5.19)$$

Note that this is very similar to the classical Mahalanobis distance which Troskie proposed to use as weights in robust regression.

$$MD_i = \sqrt{(x_i - \bar{x})^T \hat{\Sigma}^{-1} (x_i - \bar{x})} \quad (5.20)$$

The only difference between the two is that the robust distance is a robust modification of the Mahalanobis distance so it will be quite interesting to compare the performance of these two measures in outlier detection. In fact Rousseeuw, (1987) recommends using the robust distance as a way of outlier detection and as a basis for converting the raw estimates into the reweighted MCD estimators.

5.4.2 Data used in the Investigation

The data used was the identical set of returns that were used for the Markowitz formulation in Chapter 2.

5.4.3 Method of Investigation

The MCD estimator is constructed by selecting an $(h \times p)$ subset matrix of \mathbf{X} that has the least classical covariance determinant. This sounds relatively straightforward but the problem is that the number of such $(h \times p)$ subset matrices of \mathbf{X} , $\binom{n}{h}$, can be considerable large. For large data, looking for all the possible covariance matrices takes such a long time even on a good computer because the computations involved are very demanding. To solve this problem, Rousseeuw, (1987) proposed a faster method, the Fast-MCD method, which is a very clever algorithm that instead of considering all $\binom{n}{h}$ different matrices, it starts with randomly selected subset matrices then it uses an iterative procedure that converges to the solution. We used this Fast-MCD method and implemented it in MATLAB 7, using routines developed by Verboven and Hubert, (2004) in their comprehensive MATLAB toolbox called LIBRA which is a MATLAB library for robust analysis. The raw and the reweighted MCD estimates of the mean and variance were computed using the fast-MCD method and the corresponding efficient frontiers were constructed.

5.4.4 Results

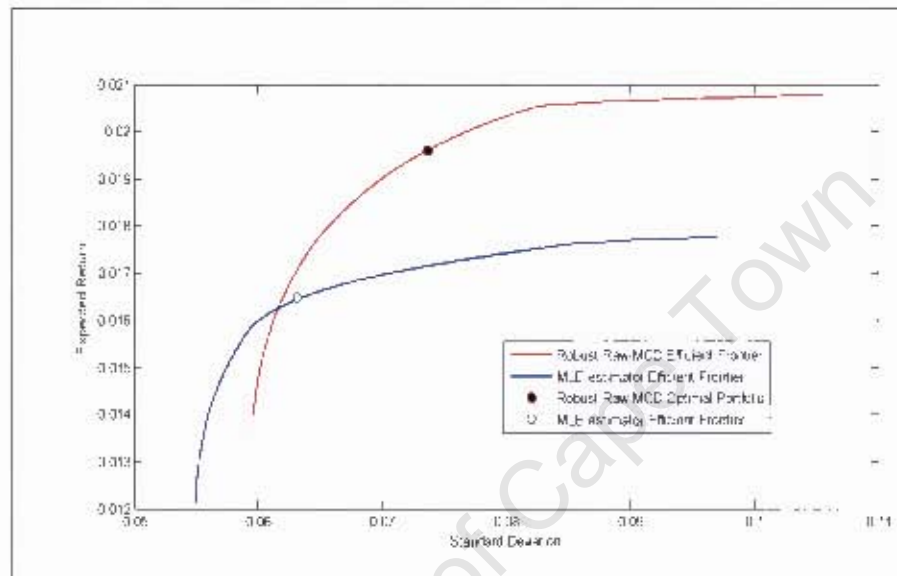


Figure 15: Robust raw-MCD Estimator Efficient Frontier

The raw-MCD estimator shown in Figure 15 shows significant differences between the classical efficient frontier and the robust frontier. For most regions of the efficient frontier, the raw-MCD estimator appears to outperform the classical MLE estimates. This is very interesting because it is telling us that if we remove the outlying observations through a robust procedure, we get higher estimates of portfolio return for most levels of risk. This must mean that most of the outlying returns must have been the negative returns since their removal resulted in higher expected returns. What is also striking is the fact that for the lower levels of risk and return, the classical efficient frontier performs better than the MCD estimate.

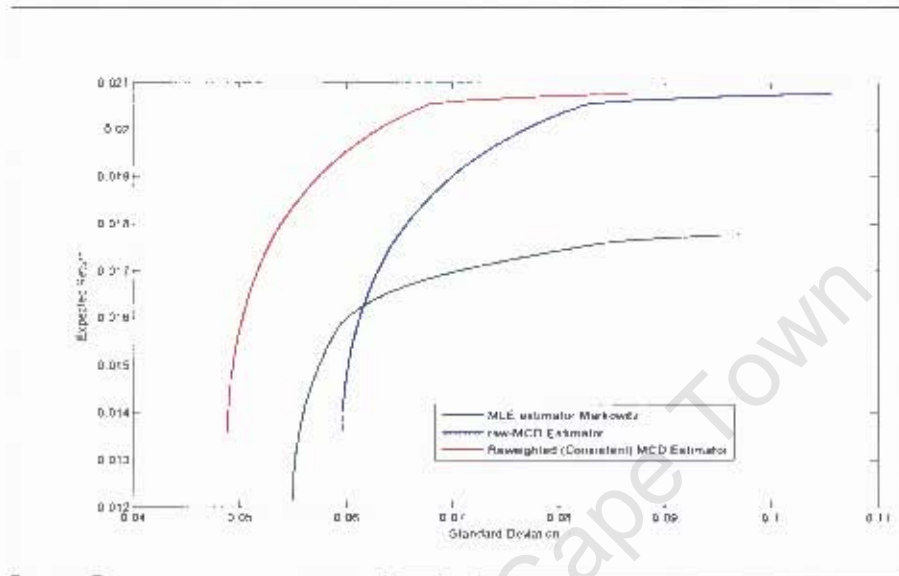


Figure 16: Reweighted MCD Estimator Efficient Frontier

Figure 16 shows the robust reweighted-MCD estimator together with the raw-MCD estimator and the classical estimator. The weights used in the reweighted-MCD estimator are shown in equation (5.16). As Rousseeuw, (1987) points out the reweighted-MCD estimate is preferred because it results in more efficient estimators. The benefit of using the reweighted-MCD estimate is that it allows us to use statistical methods to identify the outliers through the robust distance measure which is a more robust modification of the Mahalanobis distance. The reweighted estimator produced an efficient frontier that outperformed both the raw-MCD and the classical MLE estimator in all regions. Although a rational investor would clearly prefer the reweighted-MCD efficient frontier we cannot conclude that it is better. The only information it tells us is that if we use robust estimators for our data we get higher estimates of returns and a lower levels of risk.

The composition of the optimal portfolios from the three methods is very

Table 16: Composition of the Robust Optimal Portfolios

Share	Markowitz	Raw-MCD	Reweighted-MCD
Firststrand	6.18%	14.61%	14.61%
Lonmin (Jse)	42.69%	3.66%	3.66%
Pick n Pay Holdings	0.53%	4.10%	4.10%
Shoprite	29.27%	42.47%	42.47%
Sasol	21.33%	35.16%	35.16%

Table 17: Risk-Ruturn properties of the optimal portfolio

Method	Expected Return	Standard Deviation
Markowitz	19.80%	21.89%
Raw-MCD	23.64%	25.53%
Reweighted-MCD	23.52%	20.97%

different. The share Lonmin which had 42% allocation with the classical estimator now only has 3.66% allocation under the MCD estimate. That is quite a major difference in the asset allocation. Although the proportions have dramatically changed, the composition of the optimal portfolio is still the same. Notice also that although the efficient frontier from the reweighted-MCD estimate is different to the raw-MCD estimate, the proportions of the assets in the optimal risky portfolio are identical. This is really surprising considering that the raw-MCD covariance matrix is quite different to the reweighted one.

A useful feature of the MCD-estimators is the flexibility that it gives the practitioner depending on the amount of contamination in the data. Keeping in mind that the proportion of outliers that the MCD estimator can resist is $(1 - \frac{h}{n})$, for $(\frac{n}{2} < h < n)$, we can achieve different levels of robustness depending on our choice of $\frac{h}{n}$. In the efficient frontiers produced above, we used $\frac{h}{n} = 0.9$ meaning that the estimator could resist at most 10% of outlying observations. What would happen if we made our estimator even

more robust by using a lower values of $\frac{h}{n}$? Using a much lower value of $\frac{h}{n}$ value would mean that a lot more points in the data set would be considered as potential outliers. This value must be chosen carefully because if it is too high, then the estimator can break down and robustness will not be achieved.

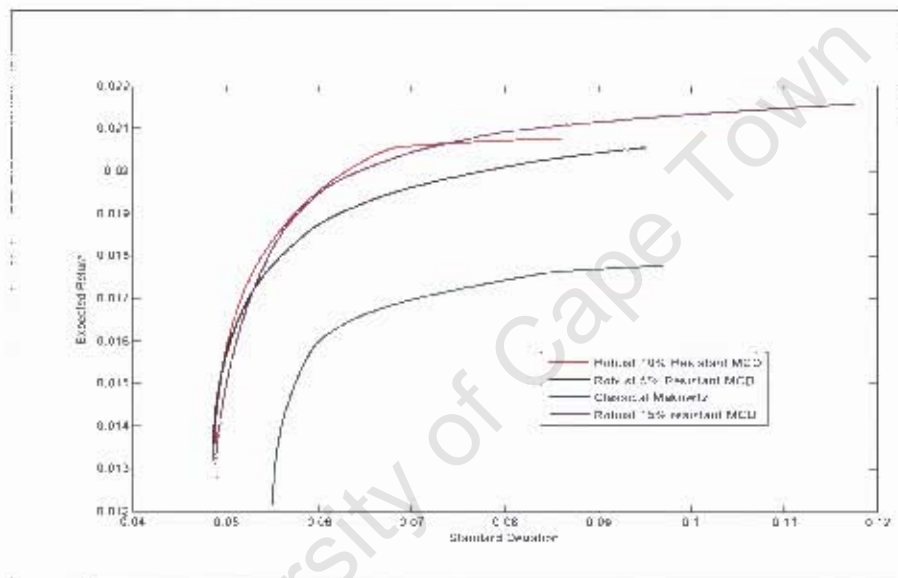


Figure 17: Robust MCD Efficient Frontiers with varying resistance

In Figure 17 the robust estimation routine using the reweighted-MCD estimator was performed for different resistance levels. The resistance level being the maximum proportion of outliers that the estimator can resist. The results show that for our data set, the different levels of resistance gave similar results. The fact that different resistance levels gave similar results is fantastic because it tells us that none of the robust estimators broke down. If one of the resistance levels had given a very different result, then it would tell us that the resistance level was not appropriate for the level of contamination in the data. It is clear that all the robust estimators produce efficient frontiers

that outperform the classical Markowitz estimator. The MCD estimators give us an additional tool of outlier detection and we can use various plots to show the presence of outliers. A useful plot is one that shows the robust distances RD_i and the Mahalanobis distances MD_i shown in figure 18

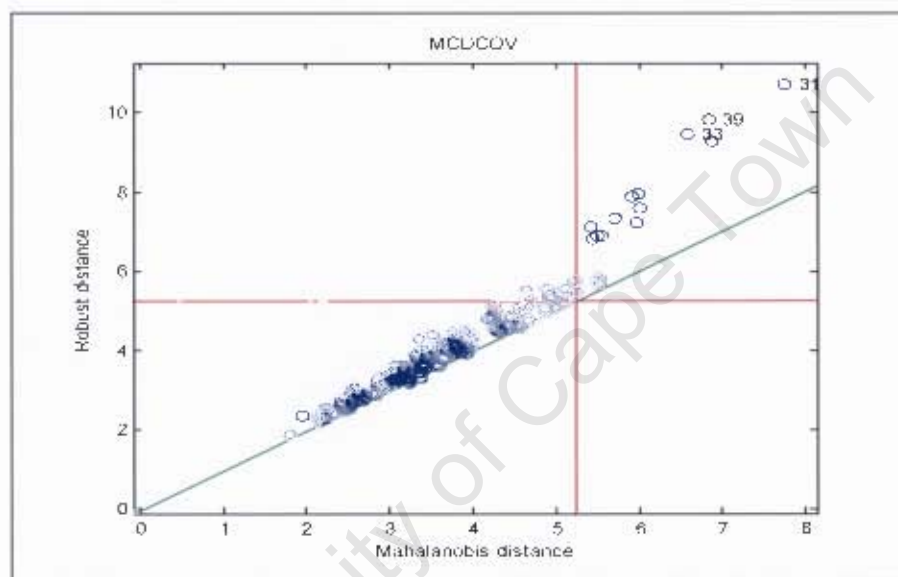


Figure 18: Robust vs Mahalanobis Distances

Figure 18 shows the robust and the Mahalanobis distance of each data point while the red lines represent the cut-off values from the $\chi_{p,0.975}^2$. The robust distance is shown to identify more outliers than the Mahalanobis distance. Every outlier according to the Mahalanobis distance is also an outlier under the Robust distance while the converse is not true. Outlier detection is very useful because once we have identified the outliers we can then use a weighting procedure that effectively deals with them. Note that in the reweighted-MCD estimator that we used, the outlying observations got a weight of zero. We may not want to use such weights if we believe that some of these outliers contain useful information. In particular, for stock

market data, outlying data points may be times when the market crashes so we really should not give them a weight of zero because we *know* that markets do crash so these outlying observations may contain this important information.

University of Cape Town

5.4.5 Conclusion

The results from the MCD estimator showed that using robust estimators as inputs in the portfolio selection process can give very different results to the classical estimators. This means that the presence of outliers in financial data should never be underestimated because of their potential impact on portfolio decisions. The efficient frontiers derived from the MCD estimators all significantly outperformed the classical estimators which tells us that the majority of the outlying observations were the low returns. The robust measure of distance that is derived from the MCD estimator managed to identify more outliers than the classical Mahalanobis distance making it a more effective tool for outlier detection when data is contaminated.

5.5 Robust Regression Techniques

In the previous section, the vector of means and covariance matrix of returns were estimated directly from the data using multivariate robust theory. We can also use robust estimators obtained from regression models. Rousseeuw, (1987) and Huber, (1981) have suggested a variety of methods for robust regression to effectively deal with outliers. The motivation for using robust estimation in regression comes from the fact that the classical least squares estimators are heavily influenced by outlying observations and so the estimates derived from that procedure can be very unreliable due to violation of the assumptions. We used regression models in the Sharpe index models in Chapter 3 using least squares estimates without worrying about the presence of outliers, so our results could be potentially inaccurate depending on the extent of contamination. The rest of this chapter will now attempt to address this problem using a variety of robust regression models.

5.5.1 Least Squares Regression

The ordinary least squares parameter estimates (OLS) of the linear regression model

$$Y_j = \beta_0 + \beta_1 X_{1j} + e_j, \quad j = 1, \dots, n. \quad (5.21)$$

are obtained if we minimize

$$\mathbf{e}'\mathbf{e} = \sum_{j=1}^n e_j^2 = \sum_{j=1}^n (Y_j - \beta_0 - \beta_1 X_{1j})^2. \quad (5.22)$$

The coefficient estimates are identical to those obtained from the maximum likelihood estimates assuming that the residuals are normally distributed, and if this is the case, then the least squares estimates are the best possible estimators. If this assumption is mildly violated in any way, then we need to identify the observations responsible for the departure from normality and deal with them either by down-weighting them or discarding them. If we are going to use the least squares estimators, we need to know exactly how sensitive our results are to these outlying observations to help us decide if the model is appropriate. One way to assess this is through the breakdown point. Recall from the previous section that the breakdown point is the highest level of contamination in the data that an estimator can function before it starts giving arbitrarily large values. As pointed out by Rousseeuw, (1987), a single observation can potentially influence the estimates due to high leverage meaning that the breakdown point of the least squares estimators is given by

$$B(T(\hat{\beta}), \mathbf{X}) = \frac{1}{n} \quad (5.23)$$

This is clearly unfavourable and it gets even worse asymptotically be-

cause the breakdown point tends to zero as the sample size increases. We thus clearly need an estimator with a higher breakdown point if we suspect contamination in the data and we detail some of the proposed estimators below. In addition, Dollinger and Staudte, (1991) showed that the influence function of the least squares estimator is unbounded, meaning that there is no guarantee that the bias will be bounded asymptotically if the data is contaminated.

5.5.2 Least Trimmed Squares Estimator

The ordinary least squares estimator (OLS) looks at the sum of the squared residuals of all the data points but since a few outliers can heavily influence the estimates we need a way of excluding these potential outliers from the minimization process. One way to do this is to minimize the sum of a subset of the smallest h squared residuals instead of all them. The Least Trimmed Squares estimator (LTS), Rousseeuw (1987), is one such estimator, defined as the solution to the minimization problem

$$\hat{\beta}_{LTS} = \min_{\beta} \sum_{i=1}^h (r^2)_{i:n} \quad (5.24)$$

with the ordered residuals $(r^2)_{1:n} \leq (r^2)_{2:n} \leq (r^2)_{3:n} \leq \dots \leq (r^2)_{n:n}$. The residuals are first squared then they are ordered and the smallest h of them are used to solve the parameters in Equation (5.24). This is a robust measure because the estimator trims the largest residuals which could potentially be indicative of outliers. The advantage of the LTS estimator is that it has a high breakdown point, in fact for a carefully chosen h , the breakdown point the LTS estimator can approach the maximum value of 50%. The LTS estimator is also asymptotically efficient which is a very desirable feature of an estimator. Equation (5.24) is not a trivial computation because

the estimator has to consider all the possible h -sized subsets and there are $\binom{n}{h}$ of them, which could be a very large number indeed. Thankfully, there are some algorithms that can compute LTS quite quickly without having to do the all $\binom{n}{h}$ computations. One such algorithm is the FAST-LTS method by Rousseeuw, (2000) which solves the LTS problem very quickly; this is the one we will use in our investigation.

5.5.3 Empirical Investigation on the Least Trimmed Squares Estimator

We will apply the LTS estimator to our data using the Sharpe single index model. The approach we take is to model the asset returns R_{it} , and the market returns I_t with the linear model,

$$R_{it} = \alpha_i + \beta_i I_t + e_{it}, \quad i = 1, \dots, p; \quad t = 1, \dots, N, \quad (5.25)$$

We have no idea what the true values of the beta coefficients are but we can estimate them from the data. We could use the least squares estimator (OLS) but we know how sensitive it is to outliers so we will use the least trimmed squares estimator given by the solution to Equation (5.24). The least trimmed squares estimator is particularly suited to modelling returns because empirical studies by Cont, (2001) have shown that return distributions have tails that are heavier than the normal distribution therefore trimming the large residuals can give us a much clearer picture of the asset dynamics. The computations will be performed in Matlab 7 using the FAST-LTS algorithm using code supplied by Verboven and Hubert, (2004). Once again, similar to the MCD estimator, the user chooses the amount of trimming that is required depending on how much contamination is suspected. In this investigation we chose to trim the worst 10% outlying observations and we then used the estimated parameters to construct the inputs to the portfolio optimization problem.

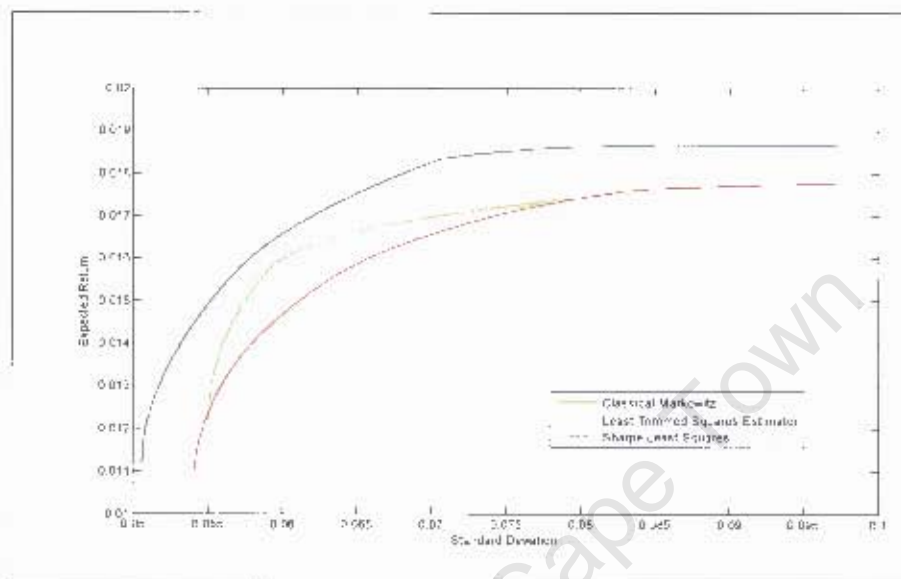


Figure 19: Least Trimmed Squares Estimator Efficient Frontiers

Figures 19 show the efficient frontiers for the Sharpe least squares estimator, the Markowitz classical estimators and the robust least trimmed estimator. The least trimmed squares estimator (LTS), has outperformed the Sharpe least squares estimator by a considerable amount. The removal of the outlying observations gives a completely different efficient frontier meaning that the data was contaminated. It is also interesting to note that the result from the least trimmed squares estimator is consistent to the results from the MCD estimator in that both methods result in similar left shifts of the efficient frontier. This is crucial because although the two approaches are different, they are both robust estimators so we expect them to have some similarities. Note however that the LTS estimator does not shift out to the left as much as the MCD estimator, probably due to the fact that we assumed that the residual errors from different shares are uncorrelated.

Table 18: OLS vs Least Trimmed Squares Estimates

Share	β_{OLS}	β_{LTS}	$\sigma_{ei(OLS)}$	$\sigma_{ei(LTS)}$
Absa	1.07	0.80	0.082	0.083
Anglos	1.30	1.30	0.058	0.057
Barloword	1.00	0.94	0.069	0.068
Bidvest	0.78	0.80	0.059	0.059
Group 5	0.72	0.80	0.128	0.128
Illovo	0.72	0.61	0.078	0.078
Firststrand	1.02	0.82	0.078	0.079
Lonmin	0.87	1.07	0.081	0.082
Mr price	0.94	0.89	0.118	0.117
Pick n pay	0.62	0.44	0.087	0.088
Shoprite	0.60	0.77	0.087	0.088
SAB	0.86	0.78	0.053	0.053
Sappi	1.15	1.09	0.092	0.092
Sasol	1.06	1.08	0.080	0.080
Standard B	0.94	0.74	0.067	0.068

Table 18 shows the regression estimates from the OLS and LTS estimators. The estimates of the beta coefficients from the LTS estimators are very different to the estimates from the ordinary least squares betas. The only exception is the estimation for Anglos which gave exactly the same estimate for beta. The difference in the estimated parameters shows how much of a difference outliers can make to the overall fit of the model. This difference will have a big impact to the portfolio selection since the optimization is very sensitive to small changes in the inputs. Interestingly though, is the fact that the residual standard errors from both the OLS and LTS estimators are almost identical. This is very surprising because it means that the estimate of the residual standard error from the ordinary least squares estimate is quite robust even if the data may contain outliers.

5.5.4 L_p Norm Estimation

The ordinary least squares estimator is actually a special case of the more general class of L_p -estimators defined as solutions to minimizing the following problem

$$L_p = \sum_{j=1}^n |Y - \beta_0 - \beta_1 X_{j1} - \cdots - \beta_p X_{jp}|^p \quad (5.26)$$

When $p = 2$ then we get the least squares solution which gives us a very nice solution for the parameter estimates. The least squares estimator is the best estimator when the residuals are normally distributed but when this assumption is violated the L_p -estimator, when $p = 1$ gives much better results. Setting $p = 1$ gives the L_1 norm referred to as the Least Absolute Deviation (LAD) estimator. The computation of the LAD is usually done using an iterative process using the assumption that we have Laplacian distributed errors. This distribution assumption is particularly suited when the residuals have much heavier tails than those from a normal distribution. Although it has the same breakdown properties as the ordinary least squares estimator, it is preferred because it is less sensitive to outliers and unlike other robust regression methods, the computations involved are not as complicated.

5.5.5 Empirical investigation on the L_1 Regression Estimator

The Sharpe single index model that is usually estimated using ordinary least squares will be estimated using the least absolute deviation (LAD) regression. The minimization problem is performed in Matlab 7 using the optimization toolbox. The optimization process was initialized using the OLS estimates as starting values for LAD minimization. The estimated parameters were then used as the inputs in the portfolio selection problem in order to construct the efficient frontier.

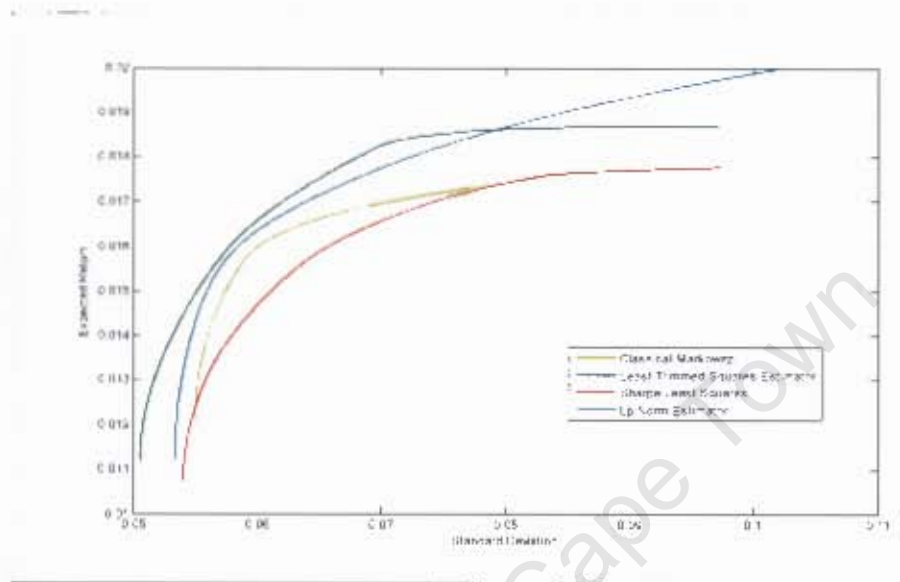


Figure 20: Lp Regression Estimator

Figure 20 shows the L_1 efficient frontier together with those from the least trimmed estimator, the least squares estimator and the classical Markowitz estimator. The efficient frontier from the L_1 shown as the blue curve has shifted significantly away from the ordinary least squares. The efficient frontier has shifted in a very similar way to how the robust least trimmed estimator and the MCD estimator moved. This shows that the L_1 estimator has responded to outliers in a similar way to other robust techniques which is encouraging because it confirms that the L_1 estimator is indeed a robust estimation procedure.

5.6 Iteratively-reweighted Least Squares Regression

5.6.1 M-estimation in Robust Regression

The most general type of robust regression estimators belong to the family of the so called *M-estimates* proposed by Huber, (1964). Given the regression model,

$$\begin{aligned} Y_i &= \beta_0 + \beta_1 X_{i1} + \cdots + \beta_p X_{ip} + e_i \\ &= \mathbf{x}'\boldsymbol{\beta} + e_i \end{aligned} \quad (5.27)$$

The *M-estimates* are derived from the minimization of the objective function

$$\sum_{i=1}^n \rho(e_i) = \sum_{i=1}^n \rho(Y_i - \mathbf{x}'_i \hat{\boldsymbol{\beta}}) \quad (5.28)$$

The function $\rho(e_i)$ is chosen appropriately to achieve robustness. Ordinary least squares estimates are actually a special case of the *M-estimates* when $\rho(e_i) = e_i^2$. According to Fox, (2002), a well chosen function $\rho(e_i)$, should have the following properties,

- $\rho(e_i) \geq 0$
- $\rho(0) = 0$
- $\rho(e_i) = \rho(-e_i)$
- $\rho(e_i) \geq \rho(e_j)$ for $|e_i| \geq |e_j|$

If $\psi = \rho'$ (the derivative of ρ) then the estimates of the parameters from the minimization of equation (5.28) are derived from differentiating it with respect to the $\boldsymbol{\beta}$. The solution then becomes

$$\sum_{i=1}^n \psi(Y_i - \mathbf{x}'_i \hat{\boldsymbol{\beta}}) \mathbf{x}'_i = 0 \quad (5.29)$$

By choosing a weighting function $w(e_i^2)$ that depends on the residuals, the ordinary least squares estimator can be improved by down-weighting the outlying observations to achieve robustness. The estimated parameters can thus be estimated from

$$\sum_{i=1}^n w_i(Y_i - \mathbf{x}'_i \hat{\boldsymbol{\beta}}) \mathbf{x}'_i = 0 \quad (5.30)$$

This gives the solution to the weighted least squares (WLS) estimator which minimizes the objective function $\sum w_i^2 e_i$. By carefully choosing the weighting function, we can penalize the outlying observations by giving them less weight so that their contribution to the sum $\sum w_i^2 e_i$, does not heavily influence the parameter estimates.

5.6.2 Robust Weighting Functions

The weighted least squares regression (WLS) requires the specification of a weighting function that can handle the presence of outlying observations. Using a weighting function in a regression setting is sometimes referred to as *bounded influence regression* because the aim is to place some sort of upper-bound on the ability of individual points to affect the overall fit of the model. The ordinary least squares estimator (OLS) have a trivial weighting function that gives an equal weighting of 1 to all observations, meaning that a single influential point can potentially affect the overall estimated parameters. Weighting observations instead of deleting them, is particularly useful for financial data because as Troskie, (2004) remarks, outliers can be due to market crashes so we cannot just delete them because they are a crucial feature of stock market dynamics. In particular, deleting outlying observations in financial data may lead to the underestimation of the variance and hence

the risk of a portfolio. A variety of weight functions have been suggested to handle outliers but we focus on the weights proposed by Huber, (1981) and Tukey, (1960).

The Huber weights can be written as

$$w_H(e_i, h_{ii}) = \frac{1}{\max(1, |k_i|)} \quad (5.31)$$

With the value of k_i derived from

$$k_i = \frac{e_i}{1.345s\sqrt{(1-h_{ii})}} \quad (5.32)$$

with e_i being the residual, h_{ii} is the leverage which is derived from the Hat matrix and s is a robust estimate of the standard deviation of the residuals. The constant value of $1.345s$ is called a tuning constant and it is chosen to ensure that the estimator is efficient which is a desirable property of an estimator. Note that the higher the value of k_i , the lower weighting that the i -th observation will have giving that particular observation less influence on the overall fit. Equation (5.32) shows that observations with large residuals will have a high value of k_i giving them less weight in the estimation. This is useful if we want to remove the outliers that may violate the regression assumptions. The residuals however are not enough to produce robust estimates because they are not very good at identifying the *influential* observations which can have a much more disastrous effect on the estimates than outliers. Indeed these influential observations can have low residuals so if the weights only depend on the residuals, we may not detect these influential observations. To measure the potential influence of an observation, the diagonal elements of the Hat matrix are used. Equation (5.32) shows that an observation with a high leverage h_{ii} , will be penalized by receiving a lower

weighting in the regression. It is for this reason that the weights in equation (5.31) have been deliberately written as a function e_i and h_{ii} , to stress the fact that the weights depend not only on the residuals, but on the leverage as well.

The Tukey Bisquare weights are defined as

$$\begin{aligned}w_{BS}(e_i, h_{ii}) &= (1 - k_i^2)^2, \text{ if } |k_i| < 1 \\ &= 0, \text{ if } |k_i| \geq 1\end{aligned}\tag{5.33}$$

With k now defined as,

$$k_i = \frac{e_i}{4.685s\sqrt{(1 - h_{ii})}}\tag{5.34}$$

The tuning constant changes from the one used in the Huber weights in order to improve efficiency.

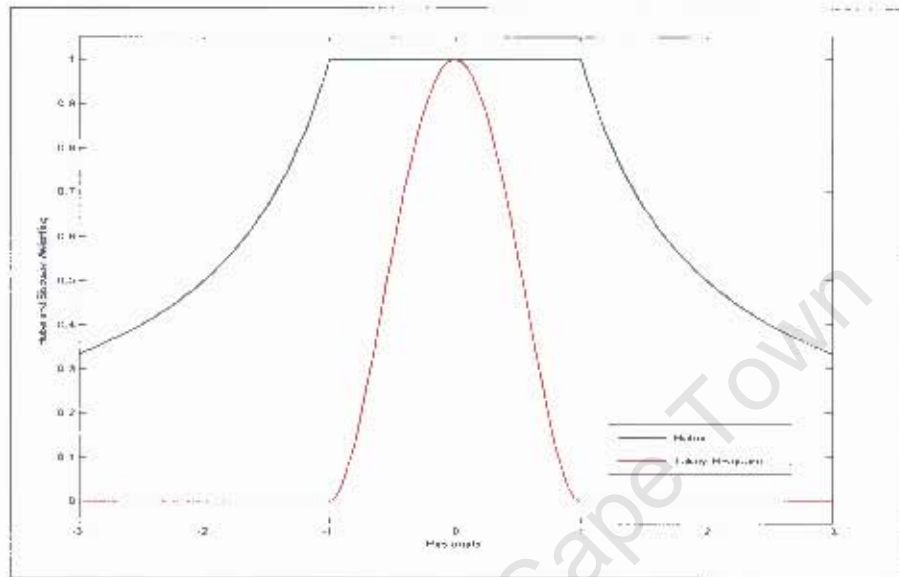


Figure 21: Huber and Tukey Bisquare Weighting Functions

Figure 21 shows the weights of the Huber and Tukey Bisquare weights plotted against the residuals. The Bisquare weights gradually decrease and they reach a point at which observations with residuals greater than a certain critical value are given no weight. The Huber weighting function is less drastic in that unlike the Bisquare function, no observation is given a weight of zero, instead the weights gradually decrease as the residuals become large. The Huber weighting function may thus be more appropriate for financial data since it down-weights outlying observations instead of completely deleting them.

5.6.3 Iteratively Reweighting Algorithm

The weights used in the weighted least squares regression depend on the residuals and the residuals depend on the estimated parameters which themselves,

depend on the weights used to estimate them. This means that fitting a set of weights can be an iterative process because once the first set of weights is used, the model can be improved by using a new set of weights that depend on the residuals from the first iteration. This process can thus be continued, each time using weights from the previous iteration until the estimated parameters converge. Fox, (2002) summarizes the iteration process in the following steps,

1. Initializes the process by finding the ordinary least squares estimates

$$\hat{\beta}^{(0)} = [\mathbf{X}'\mathbf{X}]^{-1} \mathbf{X}'\mathbf{Y} \quad (5.36)$$

2. At the k -th iteration, use the residuals from the previous iteration, $e_i^{(k-1)}$ to calculate the associated weights $w_i^{(k-1)} = w(e_i^{(k-1)})$
3. Solve for the new parameters using the diagonal matrix of weights $\mathbf{W}^{(k-1)} = \text{diag}\{w_i^{(k-1)}\}$

$$\hat{\beta}^{(k)} = [\mathbf{X}'\mathbf{W}^{(k-1)}\mathbf{X}]^{-1} \mathbf{X}'\mathbf{W}^{(k-1)}\mathbf{Y} \quad (5.37)$$

4. Go back to step 1, and continue the process until the parameters converge.

5.7 Empirical Investigation on Iteratively-Reweighted Least Squares Estimator

5.7.1 Objectives of the Investigation

We would like to investigate the effect of using weighted least squares in estimating the regression coefficients of the Sharpe single index model. The aim is to use the Huber weighting function in an iteratively reweighted least squares estimation and contrast the results to other robust regression methods such as the least trimmed squares (LTS). Keep in mind that the LTS estimator deletes outlying observations while the Huber method down-weights them, we would like to see what difference this makes to the regression estimates and the resulting efficient frontiers.

5.7.2 Method of Investigation

Using a similar data set to the one used for the Sharpe single index model, the regression coefficients from the single index model will be estimated using weighted least squares. The Huber weights will be used and the estimation will be done by iteratively re-weighting the observations until the coefficients converge. The computations will be performed in Matlab 7 and the regression will be performed for all the 15 shares. The regression estimates will be used to construct the inputs for the portfolio selection problem and the efficient frontier will be constructed.

5.7.3 Results

Table 19: OLS vs Huber Iterative Least Squares Estimates

Share	β_{OLS}	β_{WLS}	$\sigma_{ei(OLS)}$	$\sigma_{ei(WLS)}$
Absa	1.07	0.90	0.082	0.074
Anglos	1.30	1.32	0.058	0.054
Barloword	1.00	0.99	0.069	0.070
Bidvest	0.78	0.79	0.059	0.059
Group 5	0.72	0.75	0.128	0.126
Illovo	0.72	0.71	0.078	0.076
Firststrand	1.02	0.94	0.078	0.076
Lonmin	0.87	0.95	0.081	0.083
Mr price	0.94	0.90	0.118	0.103
Pick n pay	0.62	0.53	0.087	0.073
Shoprite	0.60	0.63	0.087	0.086
SAB	0.86	0.84	0.053	0.051
Sappi	1.15	1.10	0.092	0.087
Sasol	1.06	1.07	0.080	0.081
Standard B	0.94	0.80	0.067	0.056

Table 19 shows the estimated parameters from the Huber iteratively reweighted least squares estimator compared to the ordinary least squares. Although some of the beta estimates from the two models are very similar, some of the shares give very different estimates to the OLS estimates. In addition, the residual risk from the Huber weights is lower than the OLS residual risk for almost all the shares and these differences will change the inputs of the portfolio selection algorithm.

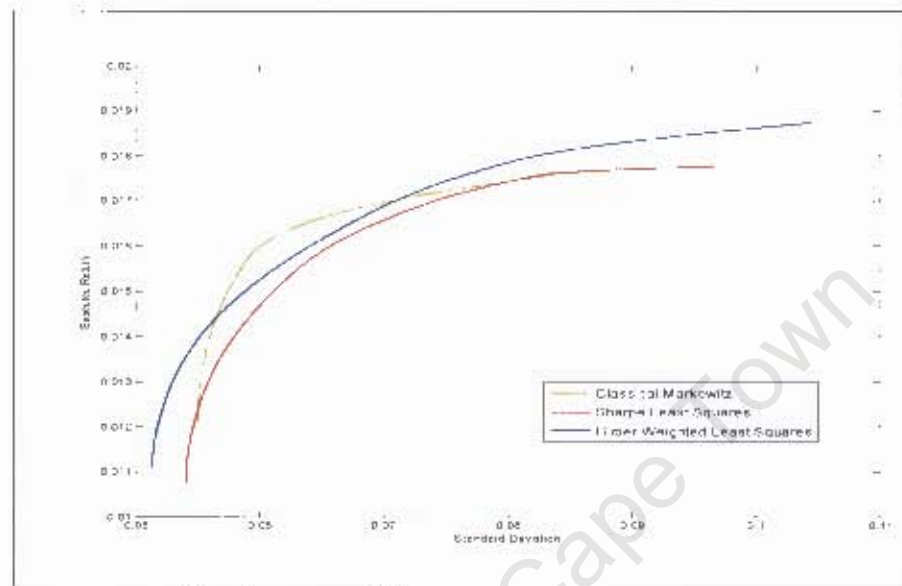


Figure 22: Huber Iteratively Reweighted Least Squares Efficient Frontier

Figure 22 shows the efficient frontier (blue line) that is produced from the Sharpe single index model using the Huber weight function in the robust iteratively reweighted least squares method. There is quite a difference in the efficient frontier produced in that the robust Huber efficient frontier shifts to the left of the OLS estimate (red line). This is consistent with the other robust estimators that were used in previous sections because they all produced a similar left shift. What is interesting is that the left shift of the Huber weighted estimator is not as severe as the shift produced by both the least trimmed squares (LTS) and the MCD estimators. This is possibly due to the fact that unlike the LTS and MCD estimators, the Huber weighting function *down-weights* the outliers instead of completely deleting them. Since most of the outlying observations were the negative returns, down-weighting them would result in lower portfolio returns than completely removing them, meaning that the resulting efficient frontier would not shift out as much as the LTS and MCD. In addition, deleting outliers would lead to a much lower

variance than down-weighting them which explains why the Huber efficient frontier does not shift out to the left as much as the LTS and MCD.

5.8 Conclusions

The investigation of robust estimation that we applied to the portfolio selection problem leads us to conclusions that can be summarized as follows.

- The classical estimates of the mean and covariance matrix used in the Markowitz model both have an unbounded influence function, meaning that these estimates can be heavily influenced by the presence of outliers.
- The MCD estimator of the mean and covariance matrix is a very robust estimator with a high breakdown point, making it a useful alternative to the classical estimators.
- The efficient frontiers and the corresponding optimal portfolios produced from the robust MCD estimator can be very different to those produced from the classical estimator. The robust estimators resulted in a left-shift of the efficient frontiers compared to the classical estimator, meaning that for the data used, the classical estimator overestimated portfolio risk or it underestimated the portfolio return.
- The MCD estimator gives a measure of the robust distance which is able to detect more outliers than the classic Mahalanobis distance. Moreover, these robust distances can be used as weights in a reweighted-MCD estimator which is a more efficient estimator than the raw estimator.
- The classical ordinary least squares estimator (OLS) is not a robust estimator because it has a very low breakdown point and it has an unbounded influence function.

- The regression can be made robust by trimming the observations with large residuals (LTS) or by using L_p methods. These robust regression methods result in significant shift in the efficient frontiers.
- Instead of deleting the outlying observations, down-weighting them is more appropriate for financial data and the resulting efficient frontiers do not shift out as severely.
- Robust estimation methods can thus lead to more relevant portfolios that are not easily influenced by outliers.

University of Cape Town

6 Principal Components Analysis

Principal components analysis is a technique that can be used to analyze multivariate data. The technique is used if a data set consists of many explanatory variables and one is interested in finding a smaller subset of the variables that can explain most of the variation. Principal components analysis can be used in a regression framework in Principal Components Regression in which orthogonal indices can be constructed from the data. This is most useful if the explanatory variables are highly collinear and as Clark and Troskie (2006) have shown, collinearity can cause the beta estimates from multiple regression to be highly unreliable. We would like to use this technique to construct optimal portfolios and then compare the results to the multiple index model results from Chapter 3. We will first present the theory involved in principal components analysis using the approach taken by Anderson (1958 p.456).

6.1 Theory of Principal Components Analysis

The key result from principal components analysis is the existence of the following orthogonal transformation, Anderson (1958 p.456)

$$\mathbf{U} = \mathbf{B}'\mathbf{X} \quad (6.1)$$

$$\text{with } \mathbf{B}'\mathbf{B} = \mathbf{I} \text{ and} \quad (6.2)$$

$$\mathbf{U}'\mathbf{U} = \mathbf{\Lambda} = \text{diag}(\lambda_i) \quad (6.3)$$

$$i = 1, \dots, p$$

\mathbf{U} is an orthogonal matrix and \mathbf{X} is the data matrix that is mean corrected with $E(\mathbf{X}) = \mathbf{0}$ and \mathbf{B} is orthogonal. The matrix $\mathbf{\Sigma}$ is the covariance of \mathbf{X} and $\mathbf{\Lambda}$ is the covariance matrix of \mathbf{U} with the diagonal elements as the eigenvalues from the solution to,

$$|\Sigma - \lambda \mathbf{I}| = 0. \quad (6.4)$$

Σ is the covariance matrix of \mathbf{X} with the property that,

$$\Sigma = \mathbf{B}\Lambda\mathbf{B}'. \quad (6.5)$$

The key quantities that need to be estimated are the eigenvalues and the corresponding eigenvectors. Anderson, (1958 p. 460) show how the maximum likelihood estimates can be computed by replacing the matrix Σ in Equation (6.4) with the maximum likelihood estimate $\hat{\Sigma}$.

6.2 Principal Components and Regression

Principal components can be used in multiple linear regression analysis especially when there is a large number of explanatory variables. Multiple linear regressions of the form

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e} \quad (6.6)$$

can give very unreliable results if the explanatory variables are highly collinear Clark and Troskie (2006). In addition, when there are a large number of explanatory variables it becomes very difficult to find a smaller set of the most important variables that account for most of the variation. Principal component regression models are a solution to the above problems because we can construct new indices from the explanatory variables that are orthogonal; hence we will not need to worry about collinearity. Remarkably, we can get a substantially smaller set of indices that explain most of the variation in the response variable which reduces the dimension of the problem. Principal components analysis gives a set of eigenvectors that we will write as

$$\mathbf{V} = (\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \dots, \mathbf{v}_p)$$

The *score vector* which is an orthogonal projection of the original data can be written as

$$\mathbf{Z} = \mathbf{XV} \quad (6.7)$$

Since we are choosing to use the first four principle components, we extract the first four \mathbf{Z} vectors ($\mathbf{z}_1, \mathbf{z}_2, \mathbf{z}_3, \mathbf{z}_4$) and compute the regression equation

$$\mathbf{Y}_t = \mathbf{c} + \beta_1 \mathbf{z}_{1t} + \dots + \beta_4 \mathbf{z}_{4t} + \mathbf{e}_t \quad (6.8)$$

The above method is called principal components regression. Hossain Troskie and Guo (2007) point out, instead of using the first couple of principal components we can also use the principal components that give the best model fit from a variable selection procedure of all the p components. This is because sometimes the score vectors associated with the smallest eigenvalues can produce the best model fit in the regression.

6.2.1 Dynamic Principal Components Regression

In dynamic principal component regression, we are interested in refining the regression assumptions of the model in order to get more accurate estimates of the regression parameters. Once we have fitted the principal components regression model, we can check if the assumptions of autocorrelation and homoskedasticity are satisfied. If any of these assumptions are violated, then we need to correct for the autocorrelation by fitting autoregressive (AR) terms or by using GARCH models to account for the heteroskedasticity.

1. An $\text{AR}(p)$ model in which we fit the autoregressive terms for the residuals in the regression.
2. An $\text{AR}(p)$ together with the $\text{GARCH}(1,1)$ model in order to take into account both the conditional heteroskedastic variance *and* the autocorrelation.

The above refinements to principle components regression mean that we can use the Gauss-Markoff theorem which states that as long as the assumptions are satisfied, the least squares estimates are the best (minimum variance) linear unbiased estimators.

University of Cape Town

6.2.2 Principal Components Kalman Filtering

Principal component regression works by constructing a set of orthogonal indices from the explanatory variables that are used in the following regression,

$$\mathbf{Y}_t = \mathbf{c} + \beta_1 \mathbf{Z}_{1t} + \dots + \beta_4 \mathbf{Z}_{4t} + \mathbf{e}_t \quad (6.9)$$

The beta coefficients are treated as constants but we propose to drop this assumption by treating them as dynamic time varying variables as shown in the following model,

$$\begin{aligned} \mathbf{Y}_t &= \mathbf{c} + \beta_{t1} \mathbf{Z}_{t1} + \dots + \beta_{t4} \mathbf{Z}_{t4} + \mathbf{e}_t \\ \beta_{ti} &= \beta_{t-1i} + \mathbf{w}_{ti} \end{aligned} \quad (6.10)$$

We can thus use the Kalman filter and treat the response variable as the signal equation and treat the beta coefficients as the state variables. This is what we propose and we will call it Principal Components Kalman Filtering. The advantage of this method is that it uses orthogonal indices derived from the principal components while at the same time using dynamic beta estimates for the betas. We intend to use this model to construct optimal portfolios from the efficient frontier.

Table 20: Explanatory Variables

Index	Variable
Jse Financial & Industrials	x3
Jse Gold Miners	x4
Jse Industrials	x5
Jse Mining	x6
Jse All Share	x7
Jse Platinum	x8
Jse excl. Resources	x9
Rand USD Exchange rate	x10

6.3 Empirical Investigation of Principal Components Analysis

6.3.1 Objectives of the Investigation

The singular value decomposition of a matrix gives us a way of constructing orthogonal indices for principal components regression. We would like to apply this method to our portfolio of shares in order to get estimates for the mean and covariance matrix of the asset returns. We are interested in investigating the effects of autocorrelation and heteroskedasticity on the efficient frontier derived from principal components regression. It will be very interesting to compare the results from these dynamic models to those from the Sharpe Multiple Index Model. Finally, we will explore the idea of using Principal Components Kalman Filtering in portfolio optimization.

6.3.2 Data used in the Investigation

The set of JSE indices that were used in the principal components regression are shown in table 20.

The monthly log returns of the indices were used to construct the \mathbf{X} matrix. The data for the individual shares and the indices spans from January 1996 to February 2008.

6.3.3 Method of Investigation

The first step in setting up the principal components regression involves finding the singular value decomposition (SVD) of the \mathbf{X} matrix. This step, which was performed in Eviews gives the eigenvalues and the corresponding eigenvectors. The score vectors were then constructed using equation (6.7) to give the \mathbf{Z} indices. These new orthogonal indices are then used as the explanatory variables for the regression. Since we started with a set of 8 indices, we had 8 different orthogonal indices to use in the regression. We could have chosen to use the indices corresponding to the highest eigenvalues but it was decided to use the method from Hossain Troskie and Guo (2007) which uses a variable selection procedure to choose the \mathbf{Z}_i indices that provide the best model fit. The score vectors that had the highest t -statistic were chosen as the explanatory variables. The following regression models were then fitted in Eviews

1. An AR(p) model in which only the autoregressive terms were fitted.
2. An Ar(p) together with the GARCH(1,1) model in order to take into account the conditional variance *and* the autocorrelation.
3. A model with the beta estimates treated as state variables using Principal Components Kalman Filtering.

6.3.4 Results- Principal Components Regression

There were a total of the 8 principal component indices that we could have used in the principal components regression, $PC1, PC2, PC3, \dots, PC8$. After a model selection procedure was performed on the full model, four of the possible 8 indices appeared to give the best overall fit for all the shares. These were $PC1, PC2, PC3$, and $PC5$.

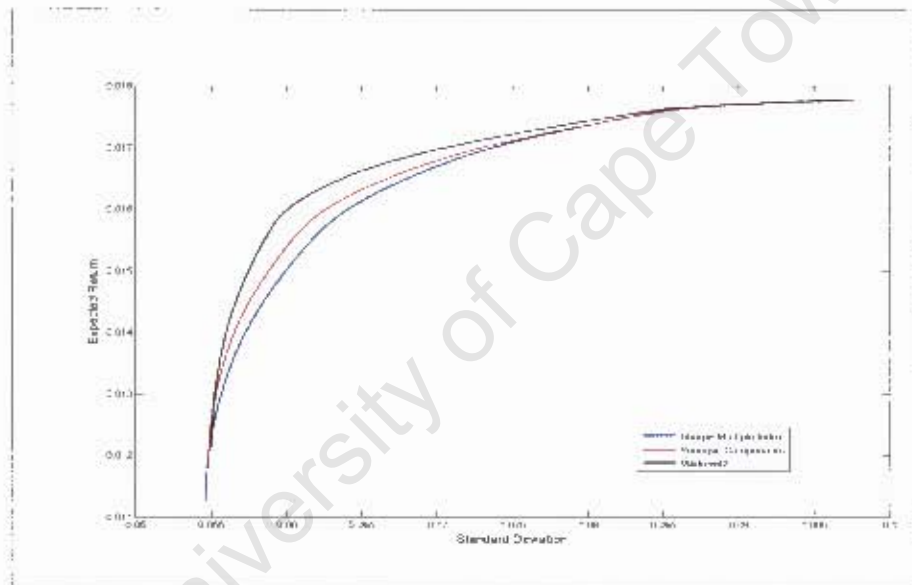


Figure 23: Principal Components Efficient Frontier

Figure 23 shows the effect of estimating the efficient frontier from principal components. The principal component regression efficient frontier has shifted to the left compared to the Sharpe multiple index model. The shift is surprisingly very similar to the shift that was produced by Hossain, Troskie and Guo, (2007) when they used different data.

Table 21: Residual Autocorrelation

Share	AR(p)	t-stat	AR(p)	t-stat	AR(p)	t-stat
Absa	AR(1)	-1.93				
Anglos	-	-				
Barlword	AR(1)	-3.74	AR(2)	-2.05		
Bidvest	AR(1)	-2.57				
Group 5	AR(1)	-2.13				
Illovo	AR(2)	-2.00				
Firstrand	-	-				
Lonmin	AR(1)	-1.58	AR(3)	-2.50		
Mr price	AR(1)	-2.25	AR(3)	-3.22	AR(6)	-2.71
Pick n pay	AR(1)	-4.68				
Shoprite	AR(5)	-2.07				
SAB	AR(1)	-2.21				
Sappi	-	-				
Sasol	-	-				
Standard B	AR(1)	-3.51				

6.3.5 Results - Dynamic Principal Components Regression

The first adjustment we make in dynamic principal components analysis is to check for autocorrelation in the residuals of the principal components regression model. The test for autocorrelation in the residuals identified the AR terms shown in table 21. It is clear that there was a significant amount of autocorrelation in the residuals from the principle components regression models. This clearly violates the regression assumptions and it means that we cannot use the Gauss-Markoff theorem meaning that the regression estimates may be unreliable. A dynamic model using the AR terms was fitted to account for the autocorrelation and the efficient frontier was computed.

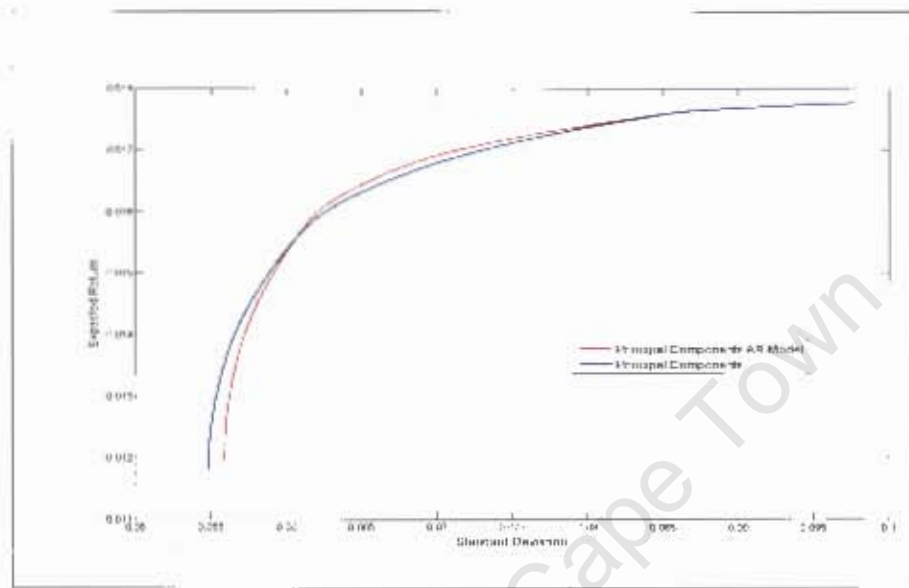


Figure 24: Principal Components AR Efficient Frontier

Figure 24 shows the efficient frontier from the AR principle components regression model together with the ordinary principle component regression model. Fitting the AR terms produces quite a different efficient frontier because it results in different estimates for the betas and the residual variance. The AR model results in an efficient frontier that has shifted to the right in the low risk regions and it has shifted to the left in the high risk region. When the principle components regression combined both the AR and GARCH(1,1) terms in a single model, the efficient frontier that was produced is shown in the Figure (25),

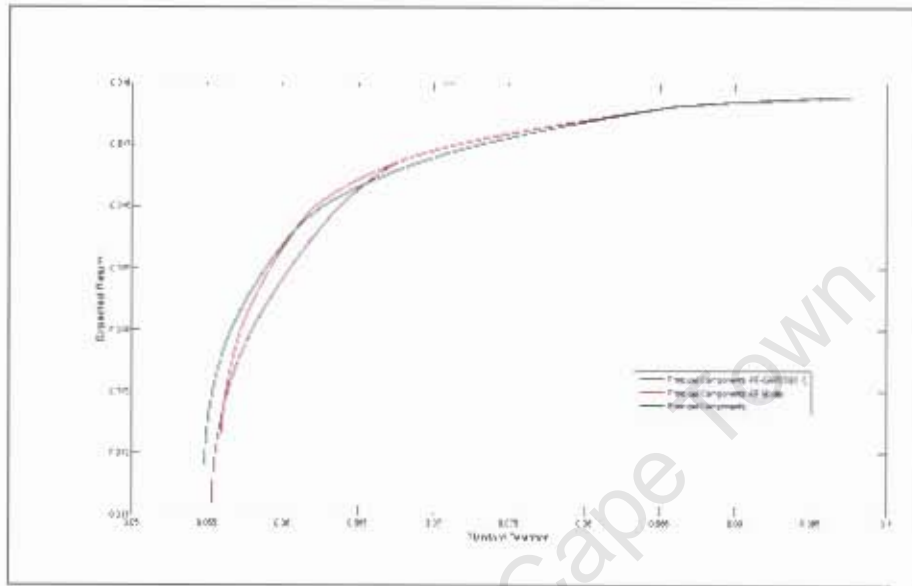


Figure 25: Combined AR-GARCH Efficient Frontier

Figure 25 shows very interesting results when the AR model is combined with the GARCH(1,1) model. Notice that if we fit just the AR terms, the efficient frontier shifts to the right (red line) in the low risk region and to the left in the high risk region. When the regression combines both the AR and the GARCH(1,1) terms, the resulting efficient frontier (purple line) has shifted significantly to the right in the low risk region and slightly to the left in the high risk region.

Table 22: R-Squared Statistics

Share	R_{OLS}^2	$R_{AR-GARCH}^2$	R_{KALMAN}^2
Absa	57%	59%	98%
Anglos	80%	81%	98%
Barlword	61%	65%	97%
Bidvest	63%	65%	96%
Group 5	17%	20%	57%
Illovo	25%	27%	78%
Firststrand	58%	57%	97%
Lonmin	55%	59%	85%
Mr price	29%	38%	50%
Pick n pay	33%	38%	85%
Shoprite	35%	36%	79%
SAB	54%	56%	97%
Sappi	39%	39%	86%
Sasol	41%	41%	84%
Standard B	64%	64%	98%

6.3.6 Results- Principal Components Kalman Filtering

Table 22 shows the R-squared statistics for the three different principle components regression models. The ordinary least squares regression model produces has the smallest R-square statistic. When the AR and GARCH(1,1) terms are fitted to the regression, there is a slight improvement in the R-squared statistic. If the principal components explanatory variables are then treated as time varying variables in the Kalman filter, there is a phenomenal increase in the R-squared statistic. The R-squared statistic reaches 98% for many of the shares. This is actually a very worrying result because it may suggest that we are now overfitting the model. Although the model has very good fit it may not be effective in terms of forecasting. This model needs much more investigation but nevertheless it is interesting to see how this affects the estimation of risk in our portfolio.

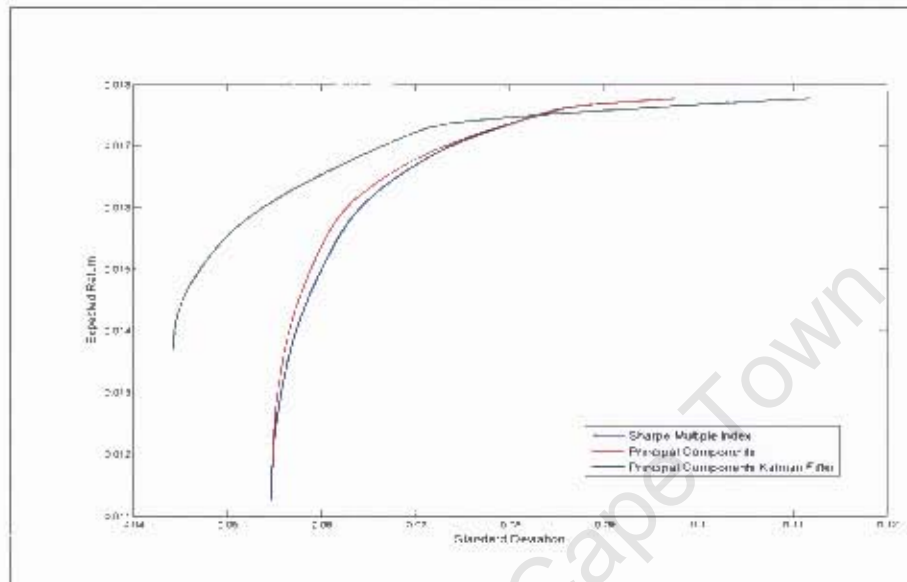


Figure 26: Principal Components Kalman Filtering

The efficient frontier produced from the Kalman filter principal components regression has shifted significantly to the left as shown by the green curve in figure 26. For most regions of the efficient frontier, the Kalman filter significantly outperforms both the Sharpe multiple index model and the least squares principal components regression model. One of the reasons for this is due to the substantial reduction in the residual variance that the Kalman filter produces.

6.4 Conclusions

The empirical investigation on the application of principal components in portfolio selection leads to the following conclusions.

- When there are a large number of possible explanatory variables or indices, principal components are able to construct a much smaller set of indices that can explain a large proportion of the total variation.
- The indices constructed from principal components are orthogonal which means that the problem of collinearity in multiple linear regression which was discussed by Clark and Troskie, (2006) was avoided.
- Correcting the principal components regression model for autocorrelation in the residuals can give a very different efficient frontier.
- A dynamic model that combines both the AR and GARCH(1,1) terms in the regression improves the fit of the model and it gives more reliable estimates because it allows us to use the Gauss-Markoff theorem.
- Principal Components Kalman Filtering provided the best fit out of all the models fitted due to the dramatic increase in the R-squared statistic. The resulting efficient frontier shifts significantly to the left because of the massive reduction in the residual variance. It will be very interesting to investigate how Principal Components Kalman Filtering performs in forecasting.

7 Conclusions

After using a variety of techniques for dynamic and robust estimation we summarize the findings from the various Chapters as follows:

Chapter 2: Markowitz Portfolio Theory

The Markowitz portfolio selection method gives a rational investor a way of selecting the optimal asset combinations based on the mean and variance of the returns. In addition, using the risk-free asset in the portfolio dramatically changes the expected returns of the portfolio. Using empirical data from the JSE, we were able to use Markowitz to construct optimal portfolios.

Chapter 3: Index Models of Asset Returns

The Sharpe single and multiple index models do not give accurate estimates of risk when the residuals are correlated. If there are negative correlations in the residuals, the Sharpe models will overestimate the portfolio risk. The Troskie-Hossain models are therefore more accurate since we found strong evidence of autocorrelation on the JSE. In addition, the combination of the GARCH and AR models in the estimation of the regression parameters makes the residuals much closer to satisfying the Gauss-Markoff theorem.

Chapter 4: State Space Models with the Kalman Filter

The dynamic estimation of the betas using the Kalman filter shows how the betas can dramatically change over time. The Kalman filter betas are very different to the least squares betas because the Kalman filter gives more weighting to the most recent observations. This has a big impact on portfolio risk because unlike the least squares estimates that assume that the beta is constant, the dynamic beta allows us to estimate market risk dynamically. We were also able to use the Kalman Filter to directly estimate returns in a dynamic model and this method resulted in the best performing portfolios

Chapter 5: Robust Portfolio Selection Techniques

We can evaluate the ability of an estimator to resist outlying observations through its breakdown point and influence function. Since the maximum likelihood estimators of μ and Σ have an unbounded influence function, they can lead to irrelevant portfolios when the data is contaminated. Robust methods give the most realistic portfolios because they are less influenced by contamination in the data. The investigation that was performed revealed that optimal portfolios constructed from robust estimators outperform those from the classical estimators

Chapter 6: Principal Components Analysis

Principal components analysis is a very effective tool for the analysis of high dimensional data. The optimal portfolios constructed from principal components regression are more efficient than the Sharpe Multiple index models. When the effects of autocorrelation and heteroskedasticity in the residuals are removed from principal components regressions using AR-GARCH models, the resulting efficient frontier leads to more accurate estimates of risk. The combination of Principal Components and Kalman filtering provides the best model fit; the R-squared statistics are so high that it needs further investigation.

7.1 Future Research

The Kalman filter and GARCH models should be combined in order to get a more complete dynamic model. One could use the Kalman filter to model the betas in the state equation and then use GARCH models for the residual variance in the signal equation. Initially, this issue was tackled in this thesis but the likelihood functions proved to be quite complicated to solve.

University of Cape Town

Appendix A

Eviews code

```
workfile rodwel 1 146  
smpl 2 146
```

```
'generating the log returns
```

```
genr r1=log(absa)-log(absa(-1))  
genr r2=log(anglos)-log(anglos(-1))  
genr r3=log(barloworld)-log(barloworld(-1))  
genr r4=log(bidvest)-log(bidvest(-1))  
genr r5=log(group5)-log(group5(-1))  
genr r6=log(illovo)-log(illovo(-1))  
genr r7=log(firststrand)-log(firststrand(-1))  
genr r8=log(lonmin)-log(lonmin(-1))  
genr r9=log(mrprice)-log(mrprice(-1))  
genr r10=log(picknpay)-log(picknpay(-1))  
genr r11=log(shoprite)-log(shoprite(-1))  
genr r12=log(sabmiller)-log(sabmiller(-1))  
genr r13=log(sappi)-log(sappi(-1))  
genr r14=log(sasol)-log(sasol(-1))  
genr r15=log(standardb)-log(standardb(-1))  
genr jse=log(jseover)-log(jseover(-1))
```

```
'generating the returns of the indeces
```

```
genr x1=log(jsebanks)-log(jsebanks(-1))  
genr x2=log(jsefin)-log(jsefin(-1))  
genr x3=log(jsefindi)-log(jsefindi(-1))  
genr x4=log(jsegold)-log(jsegold(-1))  
genr x5=log(jseindustrials)-log(jseindustrials(-1))  
genr x6=log(jsemining)-log(jsemining(-1))  
genr x7=jse  
genr x8=log(jseplatinum)-log(jseplatinum(-1))  
genr x9=log(jsexresources)-log(jsexresources(-1))  
genr x10=log(rand)-log(rand(-1))  
genr x11=log(oil)-log(oil(-1))
```

```
scalar n  
n=145
```

```
group marko r1 r2 r3 r4 r5 r6 r7 r8 r9 r10 r11 r12 r13 r14 r15
```

```
vector(15) marko_mean
```

```
for %k %y 1 r1 2 r2 3 r3 4 r4 5 r5 6 r6 7 r7 8 r8 9 r9 10 r10 11 r11 12 r12 13 r13 14 r14 15 r15  
marko_mean(%k) = @mean({%y})
```

```
next
```

```
matrix marko_cov = @cov(marko)*n/(n-1)
```

```

scalar rm
    scalar vm
        rm=@mean(jse)
        vm=@var(jse)*n/(n-1)
    ,
    ,
    ,
    ""
'now we do single index model
for %y r1 r2 r3 r4 r5 r6 r7 r8 r9 r10 r11 r12 r13 r14 r15
    equation {%y}_ls.ls {%y} c jse
    {%y}_ls.makesresid {%y}_residuals
    stom({%y}_residuals,{%y}_residvector)

next

group trosres r1_residuals r2_residuals r3_residuals r4_residuals r5_residuals r6_residuals r7_residuals
r8_residuals r9_residuals r10_residuals r11_residuals r12_residuals r13_residuals r14_residuals
r15_residuals

matrix residcor = @cor(trosres)

vector(15) beta
vector(15) beta_tstat
vector(15) beta_serr
vector (15) sereg
vector(15) alpha
vector(15) adjusted_R
vector(15) Rsquare

for %k %y 1 r1 2 r2 3 r3 4 r4 5 r5 6 r6 7 r7 8 r8 9 r9 10 r10 11 r11 12 r12 13 r13 14 r14 15 r15
    alpha(%k)={%y}_ls.c(1)
    beta(%k)={%y}_ls.c(2)

beta_tstat(%k)={%y}_ls.@tstats(2)
beta_serr(%k)={%y}_ls.@stderrs(2)
sereg(%k)={%y}_ls.@se
adjusted_R(%k)={%y}_ls.@rbar2
Rsquare(%k)={%y}_ls.@r2

next

vector(15) sharp_mean

for !i=1 to 15
    sharp_mean(!i)=alpha(!i)+beta(!i)*rm
next !i

matrix(15,15) phi

```

```
phi=beta*@transpose(beta)*vm
```

```
matrix (15, 15) phi2  
matrix(15,145) e 'e is the residual matrix E
```

```
for %k %y 1 r1 2 r2 3 r3 4 r4 5 r5 6 r6 7 r7 8 r8 9 r9 10 r10 11 r11 12 r12 13 r13 14 r14 15 r15
```

```
    phi2(%k,%k)={%y}_ls.@se^2  
    rowplace(e,@transpose({%y}_residvector},{%k})
```

```
next
```

```
matrix(15,15) phi3  
phi3=e*@transpose(e)/(n-2)  
residual covariance matrix
```

'tros hos single index

```
matrix(15,15) sharp_covariance  
    sharp_covariance=phi+phi2
```

```
matrix(15,15) troshos  
troshos=phi+phi3
```

```
;  
;  
;
```

```
'now we do the multiple index model
```

```
for %y r1 r2 r3 r4 r5 r6 r7 r8 r9 r10 r11 r12 r13 r14 r15  
equation {%y}_multiple.ls {%y} c x4 x7 x10  
{%y}_multiple.makesresid {%y}_multi_resids  
stom({%y}_multi_resids,{%y}_mresidv)
```

```
next
```

```
group mtrosres r1_multi_resids r2_multi_resids r3_multi_resids r4_multi_resids r5_multi_resids  
r6_multi_resids r7_multi_resids r8_multi_resids r9_multi_resids r10_multi_resids r11_multi_resids  
r12_multi_resids r13_multi_resids r14_multi_resids r15_multi_resids
```

```
matrix mresidcor = @cor(mtrosres)
```

'correlation of multiple index model residuals

```
vector(3) indexmean
the indeces
```

'mean of each of

```
indexmean(1)=@mean(x4)
indexmean(2)=@mean(x7)
indexmean(3)=@mean(x10)
```

```
group indices x4 x7 x10
matrix index_cov=@cov(indices)*n/(n-1) 'covariance matrix of index
```

```
'collecting all the betas for each share
for %y r1 r2 r3 r4 r5 r6 r7 r8 r9 r10 r11 r12 r13 r14 r15
vector(3) beta_{%y}
for each stock
vector(3) tstats_{%y}
```

'vector of betas

```
beta_{%y}(1)= {%y}_multiple.c(2)
beta_{%y}(2)= {%y}_multiple.c(3)
beta_{%y}(3)= {%y}_multiple.c(4)
```

```
tstats_{%y}(1)= {%y}_multiple.@tstats(2)
tstats_{%y}(2)= {%y}_multiple.@tstats(3)
tstats_{%y}(3)= {%y}_multiple.@tstats(4)
```

next

```
vector(15) alpha_mult
matrix(15,15) phi2_multi
(diagonal for Sharpe)
matrix(15,3) beta_mult
matrix(15,145) e_multi
index
matrix(15,3) beta_mtstat
```

'the constant of each regression

'the second term of the covariance

'the p X m matrix of betas

'e_multi is the residual matrix E for multiple

'the pXm matrix of tstats

```
for %k %y 1 r1 2 r2 3 r3 4 r4 5 r5 6 r6 7 r7 8 r8 9 r9 10 r10 11 r11 12 r12 13 r13 14 r14 15 r15
```

```
alpha_mult(%k)= {%y}_multiple.c(1)
phi2_multi(%k,%k)={%y}_multiple.@se^2
rowplace(e_multi,@transpose({%y}_mresidv},{%k}))
```

```
rowplace(beta_mult,@transpose(beta_{%y} ),%k)
rowplace(beta_mtstat,@transpose(tstats_{%y} ),%k)
```

next

```
vector(15) sharpm_mean
matrix(15,15) sharpm_cov
index
```

'identical to the sample mean

'the covariance matrix under sharpe multiple

```
sharpm_mean=alpha_mult+beta_mult*indexmean
sharpm_cov=beta_mult*index_cov*@transpose(beta_mult)+phi2_multi
```

```
matrix(15,15) phi3_multi
phi3_multi=e_multi*@transpose(e_multi)/(n-4)
```

```
matrix(15,15) troshos_multi
troshos_multi=beta_mult*index_cov*@transpose(beta_mult)+phi3_multi
```

```
'now we do state space model
```

```
for %y r1 r2 r3 r4 r5 r6 r7 r8 r9 r10 r11 r12 r13 r14 r15
```

```
delete {%y}_ss
sspace {%y}_ss
{%y}_ss.append {%y}=c(1) + sv1*jse
{%y}_ss.append @state sv1= sv1(-1)
{%y}_ss.ml
{%y}_ss.makesresid {%y}_ssresids
```

```
next
```

```
vector(15) beta_ss
vector(15) alpha_ss
vector(15) rsquare_ss
```

```
for %k %y 1 r1 2 r2 3 r3 4 r4 5 r5 6 r6 7 r7 8 r8 9 r9 10 r10 11 r11 12 r12 13 r13 14 r14 15 r15
```

```
beta_ss(%k)={%y}_ss.@coefs(4)
alpha_ss(%k)={%y}_ss.@coefs(1)
rsquare_ss(%k)={%y}_ss.@r2(1)
```

```
next
```

```
vector(15) sharpss_mean
```

```
for !i=1 to 15
    sharpss_mean(!i)=alpha_ss(!i)+beta_ss(!i)*rm
next !i
```

```
matrix (15,15) phi2_ss
matrix(15,145) e_multi
vector(15) ss_se
```

matrix E 'e_multi is the residual

for %k %y 1 r1 2 r2 3 r3 4 r4 5 r5 6 r6 7 r7 8 r8 9 r9 10 r10 11 r11 12 r12 13 r13 14 r14 15 r15

```
    phi2_ss(%k,%k)={%y}_ss.@se(1)^2
ss_se(%k)={%y}_ss.@se(1)
```

next

matrix(15,15) phi_ss

```
phi_ss=beta_ss*@transpose(beta_ss)*vm
```

```
matrix(15,15) sharpss_cov
    sharpss_cov=phi_ss+phi2_ss
```

‘
‘
‘

'now we multiple index ss model

for %y r1 r2 r3 r4 r5 r6 r7 r8 r9 r10 r11 r12 r13 r14 r15

```
delete {%y}_ssmult
sspace {%y}_ssmult
{%y}_ssmult.append {%y}=c(1) + sv1*x4+sv2*jse+sv3*x10
```

```
{%y}_ssmult.append @state sv1=sv1(-1)
{%y}_ssmult.append @state sv2= sv2(-1)
{%y}_ssmult.append @state sv3= sv3(-1)
```

```
{%y}_ssmult.ml
delete {%y}_ssresids
next
```

'collecting all the betas for each share

```
for %y r1 r2 r3 r4 r5 r6 r7 r8 r9 r10 r11 r12 r13 r14 r15
vector(3) beta_ss_{%y}
for each stock
```

'vector of betas

```
beta_ss_{%y}(1)= {%y}_ssmult.@coefs(6)
beta_ss_{%y}(2)= {%y}_ssmult.@coefs(7)
beta_ss_{%y}(3)= {%y}_ssmult.@coefs(8)
```

next

```
vector(15) alpha_ssmult
matrix(15,15) phi2_ssmult
```

```

matrix(15,3) beta_ssmult                                     'the p X m matrix of betas

for %k %y 1 r1 2 r2 3 r3 4 r4 5 r5 6 r6 7 r7 8 r8 9 r9 10 r10 11 r11 12 r12 13 r13 14 r14 15 r15

alpha_ssmult(%k)={%y}_ssmult.@coefs(1)
phi2_ssmult(%k,%k)={%y}_ssmult.@se(1)^2

rowplace(beta_ssmult,@transpose(beta_ss_{%y} ),%k)

next

for %y r1 r2 r3 r4 r5 r6 r7 r8 r9 r10 r11 r12 r13 r14 r15
delete beta_ss_{%y}
delete beta_{%y}

next

vector(15) sharpssm_mean
matrix(15,15) sharpssm_cov                                 'the covariance matrix under sharpe
multiple index

sharpssm_mean=alpha_ssmult+beta_ssmult*indexmean
sharpssm_cov=beta_ssmult*index_cov*@transpose(beta_ssmult)+phi2_ssmult
,
,
,

'now we do principal components
group pcax  x3 x4 x5 x6 x7 x8 x9 x10

matrix w
matrix v

matrix xmat=@convert(pcax)
matrix u=@svd(xmat,w,v)

matrix z=xmat*v

vector z1=@columnextract(z,1)
vector z2=@columnextract(z,2)
vector z3=@columnextract(z,3)
vector z4=@columnextract(z,4)
vector z5=@columnextract(z,5)
vector z6=@columnextract(z,6)
vector z7=@columnextract(z,7)
vector z8=@columnextract(z,8)

mtos(z1,z1s)
mtos(z2,z2s)
mtos(z3,z3s)
mtos(z4,z4s)
mtos(z5,z5s)
mtos(z6,z6s)
mtos(z7,z7s)

for %k %y 1 r1 2 r2 3 r3 4 r4 5 r5 6 r6 7 r7 8 r8 9 r9 10 r10 11 r11 12 r12 13 r13 14 r14 15 r15

equation r{%k}zpc.ls {%y} c z1s z2s z3s z5s

```

next

vector(4) zsmmeans
the pc indeces

'mean of each of

zsmmeans(1)=@mean(z1s)
zsmmeans(2)=@mean(z2s)
zsmmeans(3)=@mean(z3s)
zsmmeans(4)=@mean(z5s)

group pcindex z1s z2s z3s z5s
matrix pcincov=@cov(pcindex)*n/(n-1)

'collecting all the index betas for each share
for %y r1 r2 r3 r4 r5 r6 r7 r8 r9 r10 r11 r12 r13 r14 r15
vector(4) betaz_{'%y}'
for each stock

'vector of betas

betaz_{'%y'}(1)= {'%y'}zpc.c(2)
betaz_{'%y'}(2)= {'%y'}zpc.c(3)
betaz_{'%y'}(3)= {'%y'}zpc.c(4)
betaz_{'%y'}(4)= {'%y'}zpc.c(5)

next

vector(15) alphaz
matrix(15,15) phi2z
matrix(15,4) betaz

'the constant of each regression
'the second term of the covariance
'the p X m matrix of betas

for %k %y 1 r1 2 r2 3 r3 4 r4 5 r5 6 r6 7 r7 8 r8 9 r9 10 r10 11 r11 12 r12 13 r13 14 r14 15 r15

alphaz(%k)= {'%y'}zpc.c(1)
phi2z(%k,%k)={'%y'}zpc.@se^2

rowplace(betaz,@transpose(betaz_{'%y'}),%k)

next

vector(15) pc_mean
matrix(15,15) pc_cov
index

'identical to the sample mean
'the covariance matrix under sharpe multiple

pc_mean=alphaz+betaz*zsmmeans
pc_cov=betaz*pcincov*@transpose(betaz)+phi2z

for %y r1 r2 r3 r4 r5 r6 r7 r8 r9 r10 r11 r12 r13 r14 r15

delete betaz_{'%y}'
next

Appendix B

MATLAB 7 code

Quadratic Optimization

```
[PortRisk, PortReturn, PortWeights]=frontcon(ExpReturn,ExpCovariance,Numports,  
[],AssetBounds,[],[]);
```

```
portalloc (PortRisk, PortReturn, PortWeights, RisklessRate, BorrowRate)
```

```
[RiskyRisk, RiskyReturn, RiskyWts] = portalloc (PortRisk, PortReturn, PortWeights,...  
RisklessRate, BorrowRate)
```

Robust Iteratively Re-Weighted Regression

```
[b1,stat1] = robustfit(jse,r1,'huber');  
[b2,stat2] = robustfit(jse,r2,'huber');  
[b3,stat3] = robustfit(jse,r3,'huber');  
[b4,stat4] = robustfit(jse,r4,'huber');  
[b5,stat5] = robustfit(jse,r5,'huber');  
[b6,stat6] = robustfit(jse,r6,'huber');  
[b7,stat7] = robustfit(jse,r7,'huber');  
[b8,stat8] = robustfit(jse,r8,'huber');  
[b9,stat9] = robustfit(jse,r9,'huber');  
[b10,stat10] = robustfit(jse,r10,'huber');  
[b11,stat11] = robustfit(jse,r11,'huber');  
[b12,stat12] = robustfit(jse,r12,'huber');  
[b13,stat13] = robustfit(jse,r13,'huber');  
[b14,stat14] = robustfit(jse,r14,'huber');  
[b15,stat15] = robustfit(jse,r15,'huber');
```

Least Absolute Deviations

```
ExpReturn=alpha+beta.*rm;  
ExpCovariance=(beta*beta').*vm+phi2
```

```
%% 15. Minimizing the sum of absolute deviations
```

```
%{
```

```
Minimizing the sums of squares of errors is appropriate when
```

the noise in your model is normally distributed. Its not uncommon to expect a normal error structure. But sometimes we choose instead to minimize the sum of absolute errors.

How do we do this? Its a linear programming trick this time. For each data point, we add a pair of unknowns called slack variables. Thus

$$y(i) = a + b*x(i) + u(i) - v(i)$$

Here the scalars a and b, and the vectors u and v are all unknowns. We will constrain both u(i) and v(i) to be non-negative. Solve the linear programming system with equality constraints as above, and the objective will be to minimize sum(u) + sum(v).

The total number of unknowns will be 2+2*n, where n is the number of data points in our "regression" problem.

```
%}
%%
%%
x = jse;
y = r15
close
plot(x,y,'o')
title 'Linear data with noise'
xlabel 'x'
ylabel 'y'

%%
% formulate the linear programming problem.
n = length(x);
% our objective sums both u and v, ignores the regression
% coefficients themselves.
f = [0 0 ones(1,2*n)];

% a and b are unconstrained, u and v vectors must be positive.
LB = [-inf -inf , zeros(1,2*n)];
% no upper bounds at all.
UB = [];
% Build the regression problem as EQUALITY constraints, when
% the slack variables are included in the problem.
Aeq = [ones(n,1), x, eye(n,n), -eye(n,n)];
beq = y;

% estimation using linprog
```

```
params = linprog(f,[],[],Aeq,beq,LB,UB);

% we can now drop the slack variables
coef = params(1:2);

% and plot the fit
plot(x,y,'o',x,coef(1) + coef(2)*x,'-')
title 'Linprog solves the sum of absolute deviations problem (1 norm)'
xlabel 'x'
ylabel 'y'
mse=((sum((y-coef(1) - coef(2)*x).^2))/143);
alpha(15)=coef(1);
beta(15)=coef(2);
phi2(15,15)=mse;
```

University of Cape Town

References

- [1] Anderson T.W. (1984). *An Introduction to Multivariate Statistical Analysis*. New York, John Wiley and Sons, Inc, New York.
- [2] Barr, D.G.I. (1981). *A contribution to Adaptive Robust Estimation*. Unpublished Ph.D. thesis, University of Cape Town.
- [3] Bowie, D.C. (1994). *Thin trading, non-normality and the estimation of systematic risk on small stock markets*. Unpublished Ph.D. thesis, University of Cape Town.
- [4] Bollerslev, T., Chou, R. Y., Kroner, K. F. (1992) *ARCH modeling in finance*. Journal of Econometrics 52 (1992) 5-59.
- [5] Burmeister, E., Wall, K., and Hamilton, J. (1986). *Estimation of Unobserved Expected Monthly Inflation Using Kalman Filtering*. Journal of Business and Economic Statistics 4, 147-160.
- [6] Clark, A.E., and Troskie, C.G. (2006). *Regression and ICOMP - A Simulation Study*. Communications in Statistics - Simulation and Computation, 35(3): 591-603.
- [7] Cont, R. (2001). *Empirical Properties of Asset Returns: Stylized Facts and Statistical Issues*, Quantitative Finance 1, 223.
- [8] Dollinger, M.B., and Staudte, R.G. (1991). *Influence Functions of Iteratively Reweighted Least Squares Estimators*, Journal of the American Statistical Association. Vol. 86, No. 415pp. 709- 716.
- [9] Dorrington, R., and MacDonald, I., (2006). *Financial Economics*, CARE, University of Cape Town Financial Economics Course class notes.

- [10] Durbin, J., and Koopman, S.J. (2001). *Time Series Analysis by State Space Methods*. Oxford University Press, New York.
- [11] Elton, E. J., and Gruber, M. J. (1995). *Modern Portfolio Theory and Investment Analysis*. John Wiley & Sons, Canada.
- [12] Engle, R.F. (1982). *Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of U.K Inflation*. *Econometrica* 50, 987-1008
- [13] Hampel, F.R., (1968). *Contribution to the Theory of Robust Estimation*. Ph.D Thesis, University of California, Berkeley, CA.
- [14] Hampel, F.R., Ronchetti, E. M., Rousseeuw, P. J., and Stahel, W. A. (1986). *Robust Statistics an approach Based on Influence functions*. New York: John Wiley.
- [15] Haines, L.M. (2008) *Personal Communication with C.G Troskie*.
- [16] Harvey, A.C. (1990). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press, NY
- [17] Holland, P. W., and Welsch, R. E. (1977). *Robust Regression Using Iteratively Reweighted Least-Squares*. *Communications in Statistics: Theory and Methods*, A6, 1977, pp. 813-827.
- [18] Hossain, N. (2006). *Accurate Portfolio Risk-Return Structure Modeling*. University of Cape Town. (Ph.D Thesis).
- [19] Hossain, N., Troskie, C.G., and Guo, R. (2005). *GARCH(1,1) Volatility Modeling in Modern Portfolio Optimization*. Proceedings of the International Workshop on Recent Advances in Stochastic Operations Research, Canmore, Alberta, Canada, August 25-26, 2005, pp 80-87. (ISBN 4-939146-03-0).

- [20] Hossain, N., Troskie, C.G., and Guo, R. (2005). *Comparisons of the Ex Post Efficient Portfolios Generated from Sharpe and Troskie-Hossain Index Models*. Proceedings of the 10th International Conference on Industrial Engineering -Theory, Applications and Practice, Clearwater, Florida, United States of America, December 4-7, 2005, pp 412-417.
- [21] Hossain, N., Troskie, C.G., and Guo, R. (2007). *An Application of Principal Components Analysis to Portfolio Optimization under the Sharpe Multiple Index Framework*. pp 25-38, Recent Advances in Engineering and Computer Science, 2007. International Association of Engineers. (ISBN: 978-988-98671-1-9).
- [22] Huber, P. J. (1964). *Robust Estimation of a Location Parameter*. Annals of Mathematical Statistics 35:73—101.
- [23] Huber, P. J. (1981). *Robust Statistics*. Wiley. New York.
- [24] Fox, J. (2002). *Robust Regression*. Appendix to An R and S-PLUS Companion to Applied Regression.
- [25] Kalman, R. (1960). *A new approach to linear filtering and prediction problems*. Journal of Basic Engineering 82: 34 – 45.
- [26] Markowitz, H. (1952). *Portfolio Selection*. Journal of Finance, 7: pp77-91.
- [27] Maybeck, P. S. (1979). *Stochastic Models, Estimation and Control*. Academic Press, New York.
- [28] Olive, D. J. (2007). *Applied Robust Statistics*. Southern Illinois University, Department of Mathematics. Course Notes (unpublished).
- [29] Pena, D., Tiao, G.C., Tsay, R.S. (2001). *A Course in Time Series Analysis*. John Wiley and Sons. New York.

- [30] Rousseeuw, P.J. (1984). *Least Median of Squares Regression*. Journal of the American Statistical Association, Vol. 79, pp. 871-881.
- [31] Rousseeuw, P.J., and Leroy, A.M. (1987). *Robust Regression and Outlier Detection*. John Wiley and Sons, NY.
- [32] Ruppert, D. (2004). *Statistics and Finance, An Introduction*. Springer. New York.
- [33] Sharpe, W. F. (1970). *Portfolio Theory and Capital Markets*. McGraw-Hill, New York.
- [34] Shumway, R.H. (1988). *Applied Statistical Time Series Analysis*. Prentice Hall. New Jersey.
- [35] Shumway, R.H., Stoffer, D.S. (2000). *Time Series Analysis and its Applications*. Springer, NY
- [36] Troskie, C.G (2002) *Multivariate Time Series Modelling*. Unpublished Lecture Notes at University of Cape Town.
- [37] Troskie, C.G (2003) *Multivariate Financial Modelling*. Unpublished Lecture Notes at the University of Cape Town.
- [38] Troskie, C.G (2004) *Multivariate Regression*. Unpublished Lecture Notes at the University of Cape Town.
- [39] Troskie, C. G., Guo, R. K., and Hossain, N. (2008). *Accurate Estimates of Risk using the Troskie-Hossain Improved Sharpe Single Index Models. Dynamic Models*. University of Cape Town Technical Report
- [40] Tsay, R.S. (2005). *Analysis of Financial Time Series*. Wiley, NY
- [41] Tukey, J.W. (1960). *A survey of sampling from contaminated distributions*. In Contributions to Probability and Statistics, eds I. Olkin, S.

Ghurye, W. Hoeffding, W. Madow and H. Mann, pp. 448–485. Stanford: Stanford University Press.

- [42] Verboven, S., Hubert, M., (2004) *LIBRA: A MATLAB Library for Robust Analysis*.
- [43] Victoria-Feser, M., (2002). *Robust Portfolio Selection*. University de Geneve Publication.
- [44] Welsh, G., Bishop, G., (2001). *An Introduction to the Kalman Filter*. Lecture Notes from University of North Carolina at Chapel Hill Department of Computer Science.
- [45] Zuo, Y. (2001), *Some Quantitative Relationships between Two Types of Finite Sample Breakdown Point*. Statistics and Probability Letters, 51, 369-375.