Use of Exponential Smoothing Technique in Estimation of Returns in a Financial Portfolio (A Case of the Matatu Public Transport Business in Kenya)

Jumba Minyoso Sandra[1, *], Joel Cheruiyot Chelule[2], Mungatu Joseph[3]

Department of Statistics and Actuarial Studies, Jomo Kenyatta University of Agriculture and Technology, Nairobi, Kenya

Email address:
jminyoso@gmail.com (J. M. Sandra)

Abstract: This study sought to develop consistent estimators for the conditional mean and conditional volatility using exponential smoothing technique and to use the estimators for the conditional mean and conditional volatility to estimate VaR and ES of a financial asset in a given financial portfolio. In particular, we take the Kenyan Matatu business as our financial portfolio and we estimate the ES of the daily returns obtained from Matatus travelling the Nairobi –Eldoret highway as provided by CLASSIC SACCO. In estimating the conditional mean and conditional volatility of the returns of our portfolio, the study explored the exponential smoothing technique, whereby exponentially decreasing weights were assigned to the returns. The study proved that the estimators for the conditional mean and conditional volatility are consistent and also that the estimators for the conditional mean and conditional volatility when conditional mean is known, are asymptotically normal. Further the study gives the estimators for the VaR and ES and proves that the VaR estimator is consistent.

Keywords: Expected Shortfall, Exponential Smoothing, Value At Risk, Conditional Mean, Conditional Volatility

1. Introduction

1.1. Background Information

In Kenya, when the demand for public transport began to outstrip its supply shortly after Independence in 1963, the government decided to tolerate and eventually regularize what had previously been informal commercial vehicles commonly known as Matatus**.** By the mid-2000s, it is estimated that half of Nairobi's population relied on Matatus to meet their public transport needs

This Matatu industry has a direct impact on over 500,000 people who are drivers, touts, or employees of PSV Sacco's and stage attendants . As businesses, Matatus are privately owned; they do not enjoy government subsidies and are required to be registered and pay taxes.

In a legal notice issued on 23rd December 2010, by the Ministry of transport, all the Matatus were to join SACCOs or limited liability companies by the end of the year 2010. By this time some of the SACCOs were effectively managing their routes, providing credit to their members for purchase of vehicles or repairs while one had established a fueling station

for its members' vehicles with Classic Sacco being one of them. Classic Sacco manages over 170 Matatus plying the Nairobi-Eldoret route. Nairobi being the capital city of Kenya is approximately 164 miles away from Eldoret making this an approximately 5 hour drive distance when using public transport.

Returns from the Matatu business are a complete and scale-free summary of the investment opportunity in the industry

Increase of financial risks in financial portfolios the Kenyan Matatu industry; for example, have underlined the need for better financial risk measures.Matatu business owners deal with the risks associated with changes in prices that can be summarized by the variances of future returns, directly, or by their relationship with relevant covariance in a portfolio context. Forecasts of future standard deviations can provide up-to-date indications of risk, which might be used to avoid unacceptable risks . The nature of financial risks in different portfolios keeps on changing with time and therefore the methods to measure them should adapt accordingly. Furthermore, these methods should be easy to understand even in complex situations. It is in this context that

quantitative risk measures have become vital in the management of risks. This study is therefore motivated by the need to know how to estimate market risks in the Matatu Industry market.

There are many types of risks i.e. market risks, credit risks and operational risks. In this study, we are focused on one type, the market risks. Market risks are risks that arise from the changes in the prices of financial assets and liabilities (or volatilities) and measured by changes in the value of open positions or in earnings.The market risks will be quantified using a statistical tool called ES. But we cannot avoid mentioning VaR when dealing with ES, since VaR precedes ES, by definition and computation.

1.2. Statement of the Problem

Matatus account for 80% of the total public transport in the country (Muriungi, 2013).The Matatu business can yield good returns for an investor if well managed. However the Matatu business experiences many challenges. The main challenge is keeping track of these returns so as to run a profitable business. In most cases, this happens due to poor monitoring of exposure to market risks by the management. It has been difficult to manage these market risks in the absence of accurate details available from regulatory authorities and related institutions. We intend to address this challenge by estimating ES of the returns as provided by the managing body Classic SACCO.

2. Literature Review

2.1. Volatility and Value at Risk

The main characteristic of any financial asset is its return. Return is typically considered to be a random variable. Its primary usage is to estimate the value of market risk. Volatility is also a key parameter for pricing financial derivatives. All modern option-pricing techniques rely on a volatility parameter for price evaluation. Volatility is also used for risk management applications and in general portfolio management.

Volatility concept and its usage in financial risk management refer to the spread of all outcomes of an uncertain variable. In finance, we are interested in the outcomes of assets returns. Volatility is associated with the sample standard deviation of returns over some period of time. Volatility is a quantified measure of market risk. Volatility is related to risk, but it is not exactly the same. Risk is the uncertainty of a negative outcome of some event (e.g. stock returns) while volatility measures a spread of outcomes. This includes positive as well as negative outcomes.

Volatility can be used in some risk management applications, such as Value at Risk (VaR). Accurate estimates of volatility are important for option pricing, portfolio analysis and risk management methodologies, such as value at risk. The observation that many financial series exhibit volatility clustering has led to the development of a great many time series methods for volatility forecasting

As a risk measure, volatility is an important concept in finance and plays a critical role in a wide range of applications such as asset pricing, portfolio management, risk management, and option valuation. . Volatility modeling is required for VaR estimation.

The main characteristic of any financial asset is its return. Return is typically considered to be a random variable. Its primary usage is to estimate the value of market risk

Value at risk (VaR) is commonly used in the financial industry to quantify risk in asset portfolios. Value-at-risk (VaR) has become a standard measure used in financial risk management due to its conceptual simplicity, computational facility, and ready applicability. Many authors suggest that VaR has several conceptual problems for example,

- VaR measures only percentiles of profit-loss distributions, and thus disregards any loss beyond the VaR level ("tail risk"),
- VaR is not coherent since it is not sub-additive.

2.2. Expected Shortfall

Expected shortfall is defined as the conditional expectation of loss given that the loss is beyond the VaR level. ES can be used to alleviate the problems present in VaR since expected shortfall considers the loss beyond the VaR level and has been proven to be sub-additive which assures its coherence as a risk measure. Expected shortfall has less of a problem in disregarding the fat tails and the tail dependence than VaR does. VaR is not a risk measure because it does not fulfill the axiom of sub—additivity

2.3. Past Reviews on Exponential Smoothing

The two most popular time series approaches are GARCH models and smoothing methods. In contrast to the statistical rigor of GARCH models, smoothing methods provide a pragmatic, ad hoc approach to volatility forecasting Exponential smoothing is a popular approach, which has been found to perform well in empirical studies. It involves the allocation of exponentially decreasing weights to past squared shocks. Another common application of exponential smoothing is inventory control, where it is used to predict the level of a time series of demand.

An advantage of exponential smoothing is that it uses declining weights which "captures the cyclical behavior of return volatility." Exponential smoothing's simple and popular approach to volatility forecasting is to estimate the variance as a simple moving average of past squared shocks. writes that this estimator has two clear weaknesses. Firstly, if volatility clusters, there is strong appeal to giving more recent information greater weighting, and, secondly, the choice of how many past periods to include in the moving average is arbitrary. Exponential smoothing is also widely used to produce forecasts for the level of a time series . An alternative is to use GARCH the down-side of which is that it adds (some) complexity. Exponential smoothing may be a good compromise between quality and complexity for Value at Risk.

The exponential estimator is asymptotically fully adaptive to unknown conditional mean functions when its asymptotic properties are established and compared with those for the local linear estimator and the exponential estimator also shows superior performance when it is applied to estimate conditional volatility functions, ensuring its non-negativity (Su. et al, 2002). shows that the noisy nature of realized volatility may benefit from smoothing techniques

At the same time Exponential smoothing and exponentially weighted moving average give one of the best results among all volatility forecasting models. These two models have the lowest RMSE using the testing set. (Ladokhin, 2009)

2.4. Concepts and Definitions

We consider the problem of forecasting a value each period, which will be exceeded with probability $(1 - \theta)$ by the current portfolio, where $\theta \in (0,1)$ is a specified confidence level associated to the Value at Risk (VaR). Let $\{X_t\}$ be real-valued and $\{F_t, -\infty < t < \infty\}$ represent information available at time t. We will assume that $\{X_t\}$ is $\{F_t\}$ measurable. Let the sequence of random variables $\{X_t, X_{t-1}, X_{t-2}, \dots\}$ taking the values in R be stationary. Here, $\{X_t\}$ can be considered as the response variable and $\{F_{t-1}\}$(the past information of $\{X_t\}$ are the predictor variables (or covariates). Specifically, we want to estimate the conditional mean and the conditional volatility of $\{X_t\}$, given the past information, $\{F_{t-1}\}$, with the assumption that the function is completely determined by $\{X_t\}$, for $\theta \in (0,1)$. So, we have our underlying process of interest of the form as in Mwita (2003).

$$X_t = \mu_t + \sigma_t e_t, t = 1,2 \qquad (2.1)$$

Where

1) μ_t is the conditional mean function of X_t given the past information, F_{t-1},
2) σ_t is the conditional volatility function of X_t given the past information, F_{t-1},
3) and e_t are the standard error terms, which we assume to be independent of F_{t-1} and normally distributed with mean 0 and variance 1.

The conditional θ −quantile of (2.1) given F_{t-1} is then given by

$VaR_{t,\theta} = \mu_t + \sigma_t F_e^{-1}(\theta), t$ or simply

$$VaR_{t,\theta} = \mu_t + \sigma_t e_\theta$$

where $F_e^{-1}(\theta)$ is the $\theta -$ quantile of e_t and $\theta \in (0,1)$. Suppose e_t are independent and identically distributed (i.i.d) standard normal random variables, then

$$X_t \sim N(\mu_t, \sigma_t^2), t = 1,2$$

Definition (Risk)
Risk is the dispersion (volatility) of unexpected outcomes, generally the value of assets or liability of interest. Alternatively, we can also define risk as the quantifiable likelihood of loss or less-than-expected returns.

Definition (Risk measure)

Consider a set Q of real-valued random variables. A function $\rho: Q \rightarrow R$ is called a risk measure if it is
(i) monotonous:

$$X \in Q, X \geq 0 \Rightarrow \rho(X) \leq 0,$$

(ii) sub-additive:

$$X, Y, X + Y \in Q \Rightarrow \rho(X + Y) \leq \rho(X) + \rho(Y)$$

(iii) positivelyhomogeneous:

$$X \in Q, h > 0, hX \in V \Rightarrow \rho(hX) = h\rho(X)$$

and
(iv) translation invariant:

$$X \in Q, a \in R \Rightarrow \rho(X + a) = \rho(X) - a$$

Definition (Value at Risk)
The Value at Risk, $VaR_{t,\theta}$, based on negative returns or losses at time t given the past information of X_t is defined by,

$$VaR_{t,6} = inf\{X_t \in R | F_{X_t} \geq \theta\} \, \theta \in (0,1)$$

$VaR_{t,\theta}$ is the conditional θ-quantile of X_t given F_{t-1}. In essence, we find $VaR_{t,\theta}$ such that

$$\Pr[X_t < VaR_{t,\theta} | F_{t-1}] = \theta \qquad (2.2)$$

$(1 - \theta)$ is the probability of extreme losses greater than the $VaR_{t,\theta}$ usually taking the values 5% or 1% corresponding to one or ten day(s) periods respectively.
Definition (Expected Shortfall)
Suppose X_t is a random variable denoting the negative returns of a given portfolio on a specified time horizon T and $VaR_{t,\theta}$ is the VaR at the $100(1 - \theta)$ percent confidence level. The Expected Shortfall is defined by the following equation:

$$ES_\theta = E[X_t | X_t \geq VaR_{t,\theta}] = \theta \qquad (2.3)$$

2.5. Exponential Smoothing (ES)

Exponential smoothing is a popular scheme used to produce a smoothed time series. This forecasting procedure was first suggested by C. C Holt in about 1958. By definition, Exponential Smoothing is simply an adjustment technique, which takes the previous period's forecast and adjusts it up or down based on what actually occurred in that period. It accomplishes this by calculating a weighted average of the two values. The formula takes the form:

$$S_t = \lambda X_{t-1} + (1 - \lambda)S_{t-1}$$

Where X_{t-1} is the yesterday's actual value; S_t is the forecasted value; λ is the weighing factor or Smoothing constant and t is the current time period.

Whereas in single moving averages, the past observations are weighted equally, exponential smoothing scheme weight past observations using exponentially decreasing weights to forecast future values. In other words, recent observations are given relatively more weight than the older observations.

2.5.1. Single Exponential Smoothing

This smoothing scheme begins by setting S_2 to X_1, where S_t stands for smoothed observations or exponentially weighted moving average (EWMA), and X stands for the original observation. The subscripts refer to the time periods $1, 2, \ldots, n$ .For $t \geq 3$, the smoothed value S_t is found by computing

$$S_t = \lambda X_{t-1} + (1 - \lambda)S_{t-1}, 0 < \lambda \leq 1, t \geq 3 \qquad (2.4)$$

There is no S_1 . The smoothed series starts with the smoothed version of the second observation. Equation (2.4) is the basic equation of Exponential Smoothing and the constant or parameter λ is called the smoothing constant.The first forecast is very important, and the smaller the value of the λ , the more important is the selection of the initial exponentially weighted moving average. The initial smoothed observation plays an important role in computing all the subsequent EWMA's.

The following are some of the methods of initialization: Setting S_2 to X_1; Setting S_2 to the target of the process; Setting S_2 to the average of the first four or five observations.

2.5.2. Double Exponential Smoothing

Double exponential smoothing uses two constants and is better at handling trends. Single exponential smoothing does not excel in following the data when there is a trend. This situation can be improved by the introduction of a second equation with a second constant, , which must be chosen in conjunction with λ .The following are the two equations associated with double exponential smoothing:

$$S_t = \lambda X_t + (1 - \lambda)(S_{t-1} + b_{t-1}), 0 \leq \lambda \leq 1, \qquad (2.5)$$

$$b_t = \gamma(S_t - S_{t-1}) + (1 - \gamma)b_{t-1}, 0 \leq \gamma \leq 1, \qquad (2.6)$$

Here, we note that the current value of the series is used to calculate its smoothed value replacement in double exponential smoothing. There are several methods to choose the initial values for S_t and b_t. S_1 is in general set to X_1 while the following are three suggestions for b_1:

$$b_1 = X_2 - X_1$$

$$b_1 = [(X_2 - X_1) + (X_3 - X_2) + (X_4 - X_3)]/3; \text{ AND/OR}$$

$$b_1 = (X_n - X_1)/n - 1$$

Smoothing equation (2.5) adjusts S_t directly for trend of the previous period, b_{t-1}, by adding it to the last smoothed, S_{t-1}. This helps to eliminate the lag and brings S_t to the appropriate base of the current value.The second smoothing equation (2.6) then updates the trend, which is expressed as the difference between the last two values. The equation is similar to the basic form of single exponential smoothing but here applied to the updating of the trend.

Why exponential smoothing?

Exponential smoothing relies on only two pieces of data (the last period's actual value and the forecasted value for the same period). Therefore, it minimizes the data storage requirements. If we use the moving averages method to forecast, we need to have M past values. This is cumbersome if there are many items for which forecasting is required. Exponential smoothing technique is very simple in concept and easy to understand and is often used on large-scale statistical forecasting problems because it is both robust and easy to apply.

3. Estimators for the Conditional Mean and Conditional Volatility

3.1. Estimator for the Conditional Mean

Refer to model (2.1). We want to estimate the conditional mean, μ_t , given past information, F_{t-1}. In improving the normal method of estimating $\hat{\mu}$, i.e., $\hat{\mu} = \frac{1}{N}\sum_{J=1}^{N} X_j$, where n is the sample size and X_j are the observations, we introduced horizons H as the time we look back to estimate the conditional mean of the returns for the day t. So as to achieve better estimate of μ_t, the horizon should be large enough. In our case, we chose to use a horizon of 250 days.We also thought about the influence of the past on the today's mean and noted that the behavior of the matatu returns several days ago should not influence the $\hat{\mu}_t$ as much as yesterday's behavior. So we weighted the returns with weights, ω_t . These weights should decrease exponentially as the returns get older. That is, recent returns are given relatively more weight than older returns. Therefore, the weights are given by

$$\omega_t = (1 - \lambda)\lambda^{t-j} \text{ for some } \lambda \in (0,1) \qquad (3.1)$$

where λ is the weighting factor or smoothing constant.It is expected that these weights should sum up to 1, but because we cannot look back an infinite time span, the sum of the $\omega_j's$ may not sum up exactly to 1. But for horizons, H, large enough, the sum is approximately equal to 1. For example, suppose we look at a horizon H of 250 days (as in our case) and a λ of 0.95, the $\omega_j's$'s sum up to 0.99997.We now define our estimator for the conditional mean. Let $X_1, \ldots, X_H, X_{H+1}, \ldots, X_n$ be a sequence of returns, where H is the size of the horizon we take at time t and n is the sample size large enough such that we can have a number of horizons that can give a vector of estimates for the conditional mean. The estimator for the conditional mean is given by

$$\hat{\mu}_t = \sum_{j=t-H+1}^{t}(1 - \lambda)\lambda^{t-j}X_j \qquad (3.2)$$

Where $\in (0,1)$, is the smoothing constant; t is the current time period; H is the time horizon.

Definition (Point estimator)

If we have a random sample $X_1, X_2, \ldots \ldots, X_n$ from a density, say $f(x; \mu)$, which is known except for μ , then a point estimator of μ is a statistic, say $t(X_1, X_2, \ldots \ldots, X_n)$, whose value is used as an estimate of μ .

Definition (Convergence)

A sequence $\{X_n\}$ of random variables are said to converge in probability to a constant b if for every > 0;

$$\lim_{n \to \infty} \Pr\{|X_n - b| > \varepsilon\} = 0$$

or alternatively, $\lim_{n \to \infty} \Pr\{|X_n - b| \leq \varepsilon\} = 1$

Convergence in probability is denoted by $X_n \overset{p}{\to} b$

Definition (Weak Consistency)

Let $T_1, T_2, \ldots \ldots, T_n$ be a sequence of point estimators of μ, where $T_n = t(X_1, X_2, \ldots \ldots, X_n)$. The sequence $\{T_n\}$ is defined to be a weakly consistent sequence of point estimators of μ if the following conditions are satisfied:

$$\lim_{n \to \infty} E(T_n) = \mu;$$

$$\lim_{n \to \infty} Var(T_n) = 0$$

Assumptions

(A1) $E(e_i) = 0$

(A2) $\sigma^2 = E(e_i^2) < \infty$

(A3) $\lim_{|i-j| \to \infty} Cov(X_i, X_j) = 0$ where i represents an observation in one horizon e.g $X_1, X_2, \ldots \ldots, X_H$ and j represents an observation in another horizon e.g $X_1, X_2, \ldots \ldots, X_{H+1}$. Some remarks on the assumptions are in order. Assumption (A2) is made to guarantee that all variables in the model have a finite variance and assumption (A3) shows that correlation between observations diminishes to zero as the distance between X_i and X_j becomes large.

Theorem 3.1(Consistency for the conditional mean estimator)

Let $X_1, X_2, \ldots \ldots, X_H, X_{H+1}, \ldots \ldots X_n$ be a random sample of returns with mean μ_t and variance σ^2_t. Then,

$$\hat{\mu}_t = \sum_{j=t-H+1}^{t} (1-\lambda)\lambda^{t-j} X_j$$

is a weak consistent estimator of μ_t.

Proof of theorem 3.1

We show that $\lim_{H \to \infty} E(\hat{\mu}_t) = \mu_t$ and $\lim_{H \to \infty} Var(\hat{\mu}_t) = 0$. We proceed as follows:

$$\hat{\mu}_t = \sum_{j=t-H+1}^{t} (1-\lambda)\lambda^{t-j} X_j ; = (1-\lambda)\lambda^t \sum_{j=t-H+1}^{t} \frac{1}{\lambda^j} X_j$$

$$E(\hat{\mu}_t) = (1-\lambda)\lambda^t \sum_{j=t-H+1}^{t} E\{\frac{X_j}{\lambda^j}\} \quad (3.3)$$

$$= (1-\lambda)\lambda^t \sum_{j=t-H+1}^{t} \frac{\mu_t}{\lambda^j} \quad (3.4)$$

$$= (1-\lambda)\lambda^t H \frac{\mu_t}{\lambda^j} \quad (3.5)$$

$$= (1-\lambda)\lambda^{t-j} H \mu_t$$

$$= (1-\lambda)\lambda^{H-1} H \mu_t \quad (3.6)$$

$$\lim_{H \to \infty} E(\hat{\mu}_t) = \mu_t \quad (3.7)$$

$$Var(\hat{\mu}_t) = (1-\lambda)^2 \lambda^{2t} \sum_{j=t-H+1}^{t} var\left(\frac{X_j}{\lambda^j}\right) - Cov\left(\frac{X_j}{\lambda^i}, \frac{X_j}{\lambda^j}\right) \quad (3.8)$$

$$= (1-\lambda)^2 \lambda^{2t} \sum_{j=t-H+1}^{t} var\left(\frac{\sigma_t^2}{\lambda^{2j}}\right) \quad (3.9)$$

$$(1-\lambda)^2 \lambda^{2t} H\left(\frac{\sigma_t^2}{\lambda^{2j}}\right) \quad (3.10)$$

$$(1-\lambda)^2 \lambda^{2(t-j)} H\sigma^2_t = (1-\lambda)^2 \lambda^{2(H-1)} H\sigma^2_t \quad (3.11)$$

$$\lim_{H \to \infty} (Var\hat{\mu}_t) = \lim_{H \to \infty} \{(1-\lambda)^2 \lambda^{2(H-1)} H\sigma^2_t\} = 0$$

which implies that $\hat{\mu}_t \overset{p}{\to} \mu_t$.

Remarks

a) Equation (3.4) is obtained from equation (3.3) by noting that $H > 30$. That is, the time horizon we have selected is large. Therefore, by the central limit theorem, the returns are normally distributed with mean μ_t and variance σ^2_t i.e. $X_j \sim N(\mu_t, \sigma_t^2)$. Hence, we have $E\left(\frac{X_j}{\lambda^j}\right) = \frac{\mu_t}{\lambda^j}$ and $Var\left(\frac{X_j}{\lambda^j}\right) = \frac{\sigma^2_t}{\lambda^{2j}}$. In a similar way, equation (3.9) is obtained from equation (3.8), under assumption *(A3)*.

b) Equation (3.5) is obtained by summing equation (3.4) over the horizon, H. Similarly, equation (3.10) is obtained from equation (3.9).

c) Equations (3.6) and (3.11) are obtained by noting that $j = t - H + 1$

d) Equation (3.7) is obtained from equation (3.6) by noting that the weights sum up to 1 as $H \to \infty$.

Definition (Asymptotic normality, see Alexander et al (1974))

A sequence of estimators $T_1^*, T_2^*, \ldots, T_n^*, \ldots$ of μ is defined to be best asymptotically normal (BAN) if and only if the following four conditions are satisfied:

(i)   The distribution of $\sqrt{n}(T_n^* - \mu)$ approaches the normal distribution with mean 0 and variance $\sigma^{*2}(\mu)$ as n approaches infinity.

(ii) For every $\varepsilon > 0$,

$$\lim_{n \to \infty} P_\mu[|T_n^* - \mu| > \varepsilon] = 0$$

(iii) Let $\{T_n\}$ be any other sequence of simple consistent estimators for which the distribution of $\sqrt{n}(T_n - \mu)$ approaches the normal distribution with mean 0 and variance $\sigma^2(\mu)$.

(iv) $\sigma^2(\mu)$ is not less than $\sigma^{*2}(\mu)$ for all μ in any open interval.

Lemma 3.1 (George, G. (1973))

Let $\{F_H\}$ be a sequence of density functions, and let F F be a density function. Let $\emptyset_H$ be the characteristic function corresponding to F_H and $\emptyset$ be the characteristic function corresponding to. Then,

(i)   If $F_H(x) \xrightarrow[H \to \infty]{} F(x)$ for all continuity points x of F, then $\emptyset_H(t) \xrightarrow[H \to \infty]{} \emptyset(t)$, for every $t \in R$.

(ii) If $\emptyset_H(t)$ converges, as $H \to \infty$, and $t \in R$, to a function $g(t)$ which is continuous at $t = 0$, then g is a characteristic function, and if F is the corresponding density function, then

$$F_H(x) \xrightarrow[H \to \infty]{} F(x)$$

for all continuity points x of F.

Lemma 3.2 (George, G. (1973))

Let F and $\{F_H\}$ be density functions such that $F_H(x) \xrightarrow[H \to \infty]{} F(x)$, $x \in R$, and let F be continuous. Then the convergence is uniform in $x \in R$.

Theorem 3.2 (Asymptotic normality for the conditional mean estimator)

Assume *(A1), (A2), (A3)* hold. Then, $\sqrt{H}(\hat{\mu}_t - \mu_t)$ is asymptotically normal with mean 0 and variance

$H(1 - \lambda)^2 \sigma^2{}_t [H\lambda^{2(H-1)} - N\lambda^{2(N-1)}]$. That is,

$\sqrt{H}(\hat{\mu}_t - \mu_t) \overset{D}{\to} N(0, V^2)$ as $H \to \infty$, where

$$V^2 = H(1 - \lambda)^2 \sigma_t{}^2 [H\lambda^{2(H-1)} - N\lambda^{2(N-1)}]$$

Proof of theorem 3.2

We apply the central limit theorem. Let

$$G_H(x) = P[\frac{\sqrt{H}(\hat{\sigma}_t - \sigma_t)}{V} \le x]]$$

and $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt$. Then, we show that

$G_H(x) \underset{H \to \infty}{\longrightarrow} \Phi(x)$ uniformly in $x \in R$. But before we proceed to the proof proper, it is in order that we make the following comments as in George, G. (1973):

(i) Let $\{a_H\}$, $\{b_H\}$ $H = 1,2, \dots n$ be two sequences of numbers. We say that $\{a_H\}$ is $0\{b_H\}$ and we write $\{a_H\} = 0\{b_H\}$, if $\frac{a_H}{b_H} \underset{H \to \infty}{\longrightarrow} 0$. For example, if

$a_H = H$ and $b_H = H^2$, then $\{a_H\} = 0\{b_H\}$, since $\frac{H}{H^2} = \frac{1}{H} \underset{H \to \infty}{\longrightarrow} 0$. Clearly, if $\{a_H\} = 0\{b_H\}$, then $a_H = b_H O(1)$. Therefore, $0(b_H) = b_H O(1)$, .

(ii) If $a_H \underset{H \to \infty}{\longrightarrow} a$, then $(1 + \frac{a_H}{H})^H \underset{H \to \infty}{\longrightarrow} e^a$.

We now begin the proof. Let g_H be the characteristic function of G_H and $\emptyset$ be the characteristic function of Φ; that is, $\emptyset(t) = e^{-t^2/2}$, $t \in R$

By lemma 3.1, it suffices to proof that $g_H(t) \underset{H \to \infty}{\longrightarrow} \emptyset(t); t \in R$. This will imply that $G_H(x) \to \Phi(x); x \in R$. Then, by lemma 3.2, the convergence will be uniform. We have

$$\frac{\sqrt{H}(\hat{\mu}_t - \mu_t)}{V} = \frac{H\hat{\mu}_t - H\mu_t}{V\sqrt{H}} = \frac{1}{\sqrt{H}} \sum_{j=1}^{H} \frac{X_j - \mu_t}{V} = \frac{1}{\sqrt{H}} \sum_{j=1}^{H} Z_j$$

where $Z_j = \frac{X_j - \mu_t}{V}; j = 1,2, \dots . H$ with $E(Z_j) = 0, Var(Z_j) = E(Z^2{}_j) = 1$. Hence, for simplicity, writing $\sum_{j=1} Z_J$ instead of $\sum_{j=1}^{H} Z_j$, when this last expression appears as a subscript, we have

$$g_H(t) = g_{(1/\sqrt{H})\sum_{j=1} Z_j(t)} = g_{\sum_{j=1} Z_j}\left(\frac{t}{\sqrt{H}}\right) = [g_{Z_t}(\frac{1}{\sqrt{H}})]^H$$

Now consider the Taylor expansion of g_{z_1} around zero up to the second order term. Then

$$g_{z_t}\left(\frac{t}{\sqrt{H}}\right) = g_{z_t}(0) + \frac{t}{\sqrt{H}} g_{z_t}{}'(0) + \frac{1}{2!}(\frac{t}{\sqrt{H}})^2 g_{z_t}{}''(0)$$
$$+ O(\frac{t^2}{H})$$

Since

$$g_{z_t}(0) = 1, g'_{z_t}(0) = iE(Z_1) = 0$$

$$g''_{z_t}(0) = i^2 E(Z^2{}_i) = -1$$

we get

$$g_{z_t}\left(\frac{t}{\sqrt{H}}\right) = 1 - \frac{t^2}{2H} + O(\frac{t^2}{H})$$

$$= 1 - \frac{t^2}{2H} + \frac{t^2}{H} O(1) = 1 - \frac{t^2}{2H}[1 - O(1)]$$

Thus

$g_H(t) = \{1 - \frac{t^2}{2H}[1 - O(1)]\}^H$ Taking limits as $H \to \infty$ we have, $g_H(t) \underset{H \to \infty}{\longrightarrow} e^{-t^2/2};$, which is the characteristic function of Φ. Hence, the prove.

3.2. Estimator for the Conditional Volatility

Refer to model (2.1). We want to estimate the conditional volatility, σ_t , of X_t, given past information, F_{t-1} . As we did in section (3.1), we introduced the horizons H. Again, we also thought about the influence of the past on the today's mean and noted that the behavior of the matatu returns several days ago should not influence the $\hat{\sigma}_t$ as much as yesterday's behavior. So we weighted the returns with the same weights, ω_j, as in equation (3.1).

We now define our estimator for the conditional volatility. Let $X_1, \dots, X_H, X_{H+1}, \dots \dots, X_n$ be a sequence of returns defined as in section (3.1). We define the estimator for conditional volatility under the following situations:

1. When the conditional mean, μ_t, is known, the estimator is given by

$$\hat{\sigma}_t{}^2 = \sum \quad (1 - \lambda) \; \lambda^{t-j} X_j{}^2 \qquad (3.12)$$

2. When the conditional mean, μ_t , is unknown, the estimator is given by

$$\hat{\sigma}_t{}^2 = \sum_{j=t-H+1}^{t} (1 - \lambda) \; \lambda^{t-j} (X_j - \hat{\mu}_t)^2 \qquad (3.13)$$

Where in both cases, $\lambda \in (0,1), t$ and H are as defined in section (3.1).

The justification for squaring the returns, X_j , is as follows: Refer to our model (2.1), that is, $X_t = \mu_t + \sigma_t e_t$. Assuming $\mu_t = 0$, we have $X_t = \sigma_t e_t$. Thus, $X^2{}_t = \sigma^2{}_t + \sigma^2{}_t(e^2{}_t - 1)$. Taking conditional expectation, we have $E(X^2{}_t | X_{t-1}) = \sigma^2{}_t + \sigma^2{}_t E(e^2{}_t - 1) = \sigma^2{}_t.$.

Theorem 3.3 (Consistency for $\hat{\sigma}_t$ when $\hat{\mu}_t$ is known)

Let $X_1, \dots, X_H, X_{H+1}, \dots \dots, X_n$ be a sequence of returns with mean μ_t and variance,

$\sigma^2{}_t.$

Then,

$$\hat{\sigma}_t = [\sum_{j=t-H+1}^{t} (1 - \lambda) \; \lambda^{t-j} X_j{}^2]^{1/2}$$

is a simple consistent estimator for σ_t.

Proof of theorem 3.3

We show that $\lim_{H \to \infty} E(\hat{\sigma}^2{}_t) = \sigma_t{}^2$ and
$$\lim_{H \to \infty} Var(\hat{\sigma}^2{}_t) = 0$$

We proceed as follows:

$$\hat{\sigma}^2{}_t = \sum_{j=t-H+1}^{t} (1-\lambda)\,\lambda^{t-j}\,X_j{}^2\,]$$

$$= (1-\lambda)\,\lambda^t \sum_{j=t-H+1}^{t} \frac{1}{\lambda^j} X_j{}^2\,]$$

$$\mathrm{E}(\hat{\sigma}^2{}_t) = (1-\lambda)\,\lambda^t \sum_{j=t-H+1}^{t} E\left(\frac{X_j{}^2}{\lambda^j}\right) \quad (3.14)$$

$$=(1-\lambda)\,\lambda^t \sum_{j=t-H+1}^{t}\left(\frac{\sigma_t{}^2}{\lambda^j}\right) \quad (3.15)$$

$$=(1-\lambda)\,\lambda^t H \frac{\sigma_t{}^2}{\lambda^j} \quad (3.16)$$

$$= (1-\lambda)\,\lambda^{t-J} H \sigma_t{}^2$$

$$= (1-\lambda)\,\lambda^{H-1} H \sigma_t{}^2 \quad (3.17)$$

$$\lim_{H\to\infty} E(\hat{\sigma}^2{}_t) = \sigma_t{}^2 \quad (3.18)$$

$$var(\hat{\sigma}^2{}_t) = (1-\lambda)^2\lambda^{2t}\sum_{j=t-H+1}^{t} var\left(\frac{X_j{}^2}{\lambda^j}\right) - Cov\left(\frac{X_j{}^2}{\lambda^j},\frac{X_j{}^2}{\lambda^j}\right) \quad (3.19)$$

$$= (1-\lambda)^2\lambda^{2t}\sum_{j=t-H+1}^{t}\left(\frac{\sigma_j{}^2}{\lambda^{2j}}\right) \quad (3.20)$$

$$= (1-\lambda)^2\lambda^{2t}H\left(\frac{\sigma_j{}^2}{\lambda^{2j}}\right) \quad (3.21)$$

$$= (1-\lambda)^2\lambda^{2(H-1)}H\sigma_j{}^2 \quad (3.22)$$

$$= \lim_{H\to\infty} var(\hat{\sigma}^2{}_t) = \lim_{H\to\infty}\{(1-\lambda)^2\lambda^{2(H-1)}H\sigma_j{}^2\} = 0$$

which implies that $\hat{\sigma}_t \xrightarrow{p} \sigma_t$

Remarks

a) Equation (3.15) is obtained from equation (3.14) by noting that $H > 30$. That is, the time horizon we have selected is large. Therefore, by the central limit theorem, the returns are normally distributed with mean μ_t and variance $\sigma^2{}_t$ i.e. $X_j \sim N(\mu_t, \sigma^2{}_t)$. Hence, we have $E\left(\frac{X_j}{\lambda^j}\right) = \frac{\mu_j}{\lambda^j}$ and $Var\left(\frac{X_j}{\lambda^j}\right) = \frac{\sigma^2{}_t}{\lambda^{2j}}$. In a similar way, equation (3.20) is obtained from equation (3.19), under assumption *(A3)*.

b) Equation (3.16) is obtained by summing equation (3.15) over the horizon, H . Similarly, equation (3.21) is obtained from equation (3.20).

c) Equations (3.17) and (3.22) are obtained by noting that .

$$j = t - H + 1$$

d) Equation (3.18) is obtained from equation (3.17) by noting that the weights sum up to 1 as $H \to \infty$.

Theorem 3.4 (Asymptotic normality for $\hat{\sigma}_t$ when $\hat{\mu}_t$ is known)

Assume *(A1), (A2), (A3)* hold. Then, $\sqrt{H}(\hat{\sigma}_t - \sigma_t)$ is asymptotically normal with mean 0 and variance

$H(1-\lambda)^2\sigma^2{}_t[H\lambda^{2(H-1)} - N\lambda^{2(N-1)}$.That is,

$\sqrt{H}(\hat{\sigma}_t - \sigma_t) \xrightarrow{D} N(0,V^2)$ as $H \to \infty$, where

$V^2 = H(1-\lambda)^2\sigma^2{}_t[H\lambda^{2(H-1)} - N\lambda^{2(N-1)}$.

Proof of theorem 3.4

We apply the central limit theorem. Let

$$G_H(x) = P[\frac{\sqrt{H}(\hat{\sigma}_t - \sigma_t)}{V} \le x]$$

And $\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-t^2/2}dt$. Then, we show that $G_H(x) \xrightarrow[H\to\infty]{} \Phi(x)$ uniformly in $x \in R$. But before we proceed to the proof proper, it is in order that we make the following comments as in prove of theorem (3.2):

I. Let $\{a_H\}$, $\{b_H\}$ $H = 1,2,\dots$ be two sequences of numbers. We say that$\{a_H\}$ is $o\{b_H\}$ and we write $\{a_H\} = o\{b_H\}$, if $\frac{a_H}{b_H} \xrightarrow[H\to\infty]{} 0$. For example, if $a_H = H$ and $b_H = H^2$, then $\{a_H\} = 0\{b_H\}$, since $\frac{H}{H^2} = \frac{1}{H} \xrightarrow[H\to\infty]{} 0$. Clearly, if $\{a_H\} = 0\{b_H\}$, then $a_H = b_H O(1)$. Therefore, $0(b_H) = b_H O(1)$, .

II. If $a_H \xrightarrow[H\to\infty]{} a$, then $(1+\frac{a_H}{H})^H \xrightarrow[H\to\infty]{} e^a$.

We now begin the proof. Let g_H be the characteristic function of G_H and $\emptyset$ be the characteristic function of Φ; that is, $\emptyset(t) = e^{-t^2/2}$,$t \in R$

By lemma 3.1, it suffices to proof that $g_H(t) \xrightarrow[H\to\infty]{} \emptyset(t); t \in R$. This will imply that $G_H(x) \to \Phi(x); x \in R$. Then, by lemma 3.2, the convergence will be uniform. We have

$$\frac{\sqrt{H}(\hat{\sigma}_t - \sigma_t)}{V} = \frac{H\hat{\sigma}_t - H\sigma_t}{V\sqrt{H}} = \frac{1}{\sqrt{H}}\sum_{j=1}^{H}\frac{X_j - \sigma_t}{V} = \frac{1}{\sqrt{H}}\sum_{j=1}^{H} Z_j$$

where $Z_j = \frac{X_j - \sigma_t}{V}; j = 1,2,\dots.H$ with $E(Z_j) = 0, Var(Z_j) = E(Z^2{}_j) = 1$. Hence, for simplicity, writing $\sum_{j=1} Z_J$ instead of $\sum_{j=1}^{H} Z_j$, when this last expression appears as a subscript, we have

$$g_H(t) = g_{(1/\sqrt{H})\sum_{j=1}Z_j(t)} = g_{\sum_{j=1}Z_j}\left(\frac{t}{\sqrt{H}}\right) = [g_{Z_t}(\frac{1}{\sqrt{H}})]^H$$

Now consider the Taylor expansion of g_{z_1} around zero up to the second order term. Then

$$g_{z_t}\left(\frac{t}{\sqrt{H}}\right) = g_{z_t}(0) + \frac{t}{\sqrt{H}}g_{z_t}{}'(0) + \frac{1}{2!}(\frac{t}{\sqrt{H}})^2 g_{z_t}{}''(0) + O(\frac{t^2}{H})$$

Since

$$g_{z_t}(0) = 1, g'_{z_t}(0) = iE(Z_1) = 0$$

$$g''_{z_t}(0) = i^2 E(Z^2{}_i) = -1$$

we get

$$g_{z_t}\left(\frac{t}{\sqrt{H}}\right) = 1 - \frac{t^2}{2H} + O(\frac{t^2}{H})$$

$$= 1 - \frac{t^2}{2H} + \frac{t^2}{H}O(1) = 1 - \frac{t^2}{2H}[1 - O(1)]$$

Thus

$g_H(t) = \{1 - \frac{t^2}{2H}[1 - O(1)]\}^H$ Taking limits as $H \to \infty$
we have, $g_H(t) \underset{H \to \infty}{\longrightarrow} e^{-t^2/2}$; , which is the characteristic
function of Φ. Hence, the prove.

Theorem 3.5 (Consistency for $\hat{\sigma}_t$ when $\hat{\mu}_t$ is unknown)

Let $X_1, \dots, X_H, X_{H+1}, \dots, X_n$ be a random sample of
returns with mean μ_t and variance, $\sigma^2{}_t$

Then,

$\hat{\sigma}_t = [\sum_{j=t-H+1}^{t}(1-\lambda)\lambda^{t-j}(X - \hat{\mu}_t)^2]^{1/2}$ is a simple
consistent estimator for σ_t.

Proof of theorem 3.5

We show that $\lim_{H \to \infty} E(\hat{\sigma}^2{}_t) = \sigma_t{}^2$ and

$$\lim_{H \to \infty} Var(\hat{\sigma}^2{}_t) = 0$$

We proceed as follows:

$$\hat{\sigma}^2{}_t = \sum_{j=t-H+1}^{t} (1-\lambda)^{t-j} (X_j - \hat{\mu}_t)^2$$

$$= \sum_{j=t-H+1}^{t} (1-\lambda)^{H-1} (X_j - \hat{\mu}_t)^2$$

$$= (1-\lambda)^{H-1} \sum_{j=t-H+1}^{t} (X_j - \hat{\mu}_t)^2$$

$$= \frac{\hat{\sigma}^2{}_t}{(1-\lambda)^{H-1}} = \sum_{j=t-H+1}^{t} (X_j - \hat{\mu}_t)^2$$

$$= \frac{\hat{\sigma}^2{}_t}{(1-\lambda)^{(H-1)}\sigma^2{}_t} = \frac{\sum_{j=t-H+1}^{t} (X_j - \hat{\mu}_t)^2}{\sigma^2{}_t}$$

has a Chi-square distribution with $(H-1)$ degrees of
freedom. Therefore,

$E\left[\frac{\hat{\sigma}^2{}_t}{(1-\lambda)^{(H-1)}\sigma^2{}_t}\right] = (H-1)$ and

$$Var\left[\frac{\hat{\sigma}^2{}_t}{(1-\lambda)^{(H-1)}\sigma^2{}_t}\right] = 2(H-1)$$

Hence,

$$E(\hat{\sigma}_t^2) = (1-\lambda)^{H-1}(H-1)\sigma_t^2 \qquad (3.23)$$

$$\lim_{H \to \infty} E(\hat{\sigma}^2{}_t) = \sigma_t{}^2$$

$$Var(\hat{\sigma}^2{}_t) = (1-\lambda)^{2(H-1)}2(H-1)\sigma_t{}^4 \qquad (3.24)$$

$$\lim_{H \to \infty} Var(\hat{\sigma}^2{}_t) = 0$$

which implies that $\hat{\sigma}^2{}_t \xrightarrow{p} \sigma_t{}^2$ and so is $\hat{\sigma}_t$. Hence, the proof.

3.3. The Smoothing Constant

The choice of the smoothing parameter is very important.
However, although the theory of selecting this parameter is
widely expanding, there is yet no one particular method that is
universally acceptable as the standard. In any case, the basic
principle underlying the choice of a smoothing parameter, is
that the chosen value should result in the minimum mean
square error.

Definition (Mean-Squared Error (MSE))

Let $T = t(X_1, X_2, \dots, X_n)$ be an estimator of μ.

$E_\mu[(T - \mu)^2]$ is defined to be the mean-squared error of the
estimator $T = t(X_1, X_2, \dots, X_n)$.

3.4. Estimators for Value at Risk and Expected Shortfall

3.4.1. Estimator for Value at Risk (VaR)

The Value-at-Risk summarizes the expected maximum loss
(or worst loss) over a target horizon at a given confidence
level θ. In our case, we use a target horizon of 250 days and a
99% confidence level. . That means the Value-at-Risk we give
is the amount of money we will at maximum lose the next day
with probability of 99%. Or in other words, the probability to
lose at the next day more than the calculated Value-at-Risk is
less than 1%.

We obtain today's returns, X_t, as $X_t = -\log(\frac{P_t}{P_{t-1}})$ where
P_t and P_{t-1} are today's and yesterday's returns respectively..
The conditional θ -quantile function for our econometric
model (2.1) is given by

$$VaR_{t,\theta} = \mu_t + \sigma_t \quad (4.1)$$

While the estimator for the conditional θ -quantile is given
by

$$\widehat{VaR}_{t,\theta} = \hat{\mu}_t + \hat{\sigma}_t e_\theta \qquad (4.2)$$

where $\hat{\mu}_t$ and $\hat{\sigma}_t$ are as defined in equations (3.2) and (3.13)
respectively.

Consistency for the estimator for Value at Risk

We show that
$\lim_{H \to \infty} E(\widehat{VaR}_{t,\theta}) = VaR_{t,\theta}$ and $\lim_{H \to \infty} Var(\widehat{VaR}_{t,\theta}) = 0$.
We proceed as follows:

First, we check the bias

$$E(\widehat{VaR}_{t,\theta}) = E(\hat{\mu}_t) + e_\theta E(\hat{\sigma}_t) \qquad (4.3)$$

$$(1-\lambda)^{H-1}H\mu_t + e_\theta(1-\lambda)^{H-1}(H-1)\sigma_t{}^2 \quad (4.4)$$

$$\lim_{H \to \infty} E(V\hat{a}R_{t,\theta}) = \mu_t + \sigma_t e_\theta \qquad (4.5)$$

$$VaR_{t,\theta} \qquad (4.6)$$

Next, we check the asymptotic variance.

$$Var(V\hat{a}R_{t,\theta}) = Var(\hat{\mu}_t) + e_\theta{}^2 Var(\hat{\sigma}_t) - 2e_\theta Cov(\hat{\mu}_t, \hat{\sigma}_t) \quad (4.7)$$

$$= ((1 - \lambda)^2 \lambda^{H-1} H \sigma_t^2) + e_\theta^2 ((1 - \lambda)^{2(H-1)} 2(H - 1) \sigma_t^4) \quad (4.8)$$

$$(1 - \lambda)^2 \lambda^{H-1} H \sigma_t^2 (1 + e_\theta^2 (1 - \lambda)^{(H-1)} \sigma_t^2) \quad (4.9)$$

$$\lim_{H \to \infty} Var(V\hat{a}R_{t,\theta}) = 0 \quad (4.10)$$

Hence, the proof.

Remarks

a) Equation (4.4) is obtained from equation (4.13) by using assumption *(A3)* and noting that

$$E(\hat{\mu}_t) = (1 - \lambda)^{H-1} H \mu_t$$

and $E(\hat{\sigma}_t) = (1 - \lambda)^{H-1}(H - 1)\sigma_t$.

b) Equation (4.5) is obtained from equation (4.4) by noting that the weights sum up to 1 as $H \to \infty$.

c) In equation (4.9), $Var(\hat{\mu}_t)$ and $Var(\hat{\sigma}_t)$ are obtained as in equations (3.11) and (3.24) respectively.

3.4.2. Estimator for the Expected Shortfall

Expected shortfall is the conditional expectation of loss given that the loss is beyond the VaR level, and measures how much one can lose on average in the states beyond the VaR level.

Suppose X_t is a random variable denoting the negative returns of a given portfolio and $VaR_{t,\theta}$ is the VaR at the θ-confidence level. The Expected Shortfall is given by

$$ES_\theta = E[X_t | X_t \geq VaR_{t,\theta}]$$

The function for expected shortfall based on our econometric model (2.1) is given by

$$ES_\theta = E[\mu_t + \sigma_t e_t | \mu_t + \sigma_t e_t \geq \mu_t + \sigma_t e_t]$$

$$= \mu_t + \sigma_t E[e_t \geq e_\theta]$$

and the estimator for the expected shortfall is given by

$$E\hat{S}_\theta = \hat{\mu}_t + \hat{\sigma}_t E[\hat{e}_t \geq \hat{e}_\theta] = \hat{\mu}_t + \hat{\sigma}_t \frac{1}{k} \sum_{i=1}^{k} \hat{e}_t I_{\{\hat{e}_t \geq \hat{e}_\theta\}}$$

where $\hat{\mu}_t$ and $\hat{\sigma}_t$ are as defined in equations (3.2) and (3.13) respectively. k is the number of negative returns that exceed $VaR_{t,\theta}$ and $I_{\{\hat{e}_t \geq \hat{e}_\theta\}}$ is an indicator function defined as

$$I_{\{\hat{e}_t \geq \hat{e}_\theta\}} = \begin{cases} 1 & if \ \hat{e}_t \geq \hat{e}_\theta \\ 0 & if \ \hat{e}_t < \hat{e}_\theta \end{cases}$$

4. Results and Discussion

Here, we present analysis of average returns from Matatu business from May 2010 to August 2014. The results are given in the figures below.
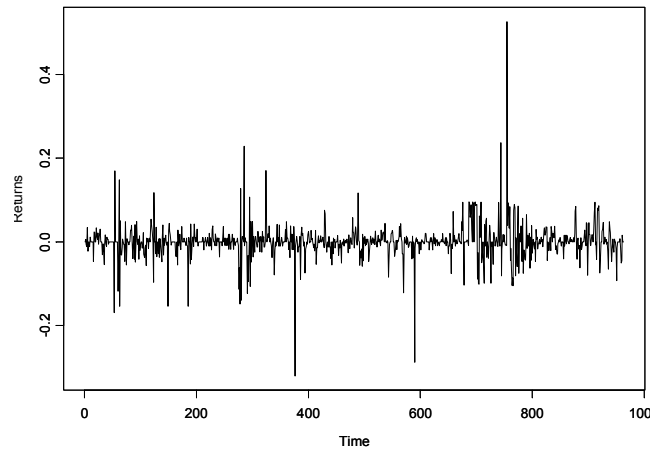

***Figure 1.** Plot of returns from May 2010 to August 2014.*

In figure 1, the upper (positive) sides are the negative returns (losses) while the lower (negative) sides are the positive returns (profits). This is because we have used the negative log returns. We observe that the returns tend to cluster as the threshold increases. This suggests that the returns are auto correlated. The clustered returns over time represent clustering of volatilities. This is supported by Engle and Manganelli (2002) who noted that the distribution of returns tend to be auto correlated. The long spikes on either side indicate extreme returns.
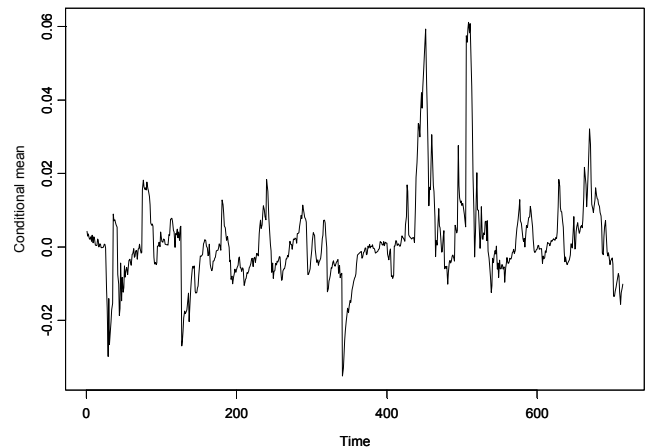

***Figure 2.** Matatu conditional means plotted against time.*

In figure 2, we can observe that the conditional mean tend to be stationary for a long period of time as it keeps on reverting to the long term value, that is, 0. There are also some few positive and negative "outliers" or extreme means that seem not be consistent with the rest of the means. The negative "outliers" suggests that the market conditions were very favorable and the positive outliers suggest that the market conditions were very unfavorable.

In figure 3, it is evident that volatility varies with time. We can also observe that large volatilities tend to be followed by large volatilities. Similarly, small volatilities tend to be followed by small volatilities. We can also see that there was a shock as indicated by the extremely large volatilities that lasted for a short period (around day 500 to 550). A shock

indicates either extreme gain or extreme loss in the market, depending on the conditions.
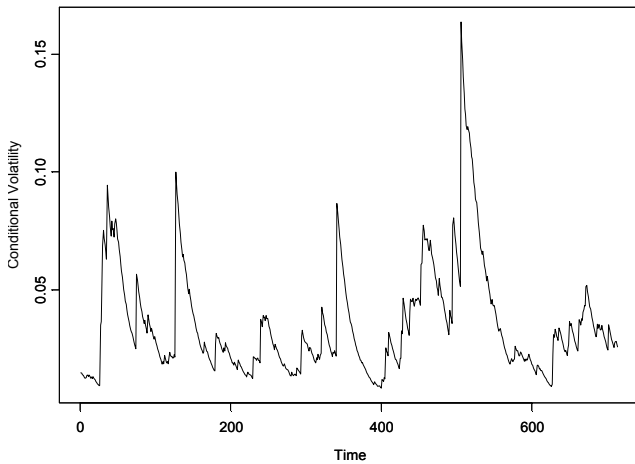

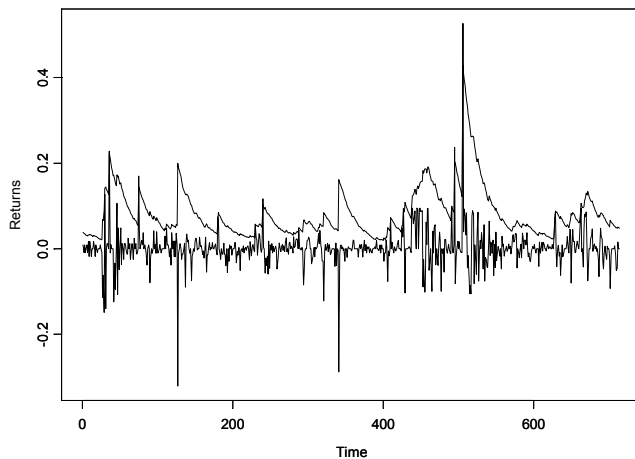*Figure 3. Matatu Conditional volatilities plotted against time.*

Figure 4. Matatu returns plotted together with $VaR_{t,0.99}$.
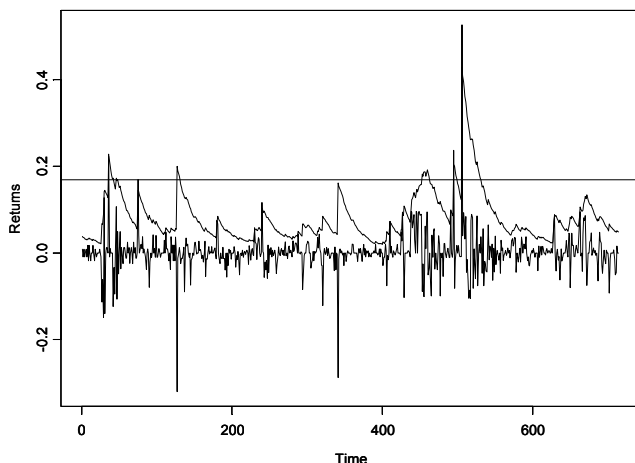

Figure 5. Plot of MATATU returns with $VaR_{t,0.99}$ and Expected Shortfall.

In figure 4, we have superimposed 0.99-conditional quantile on the returns. VaR is always based on negative returns hence the plot of $VaR_{t,0.99}$ on the upper side. From the definition of VaR, 0.99-conditional quantile means that

there is one chance in a hundred chances, under normal market conditions, for a loss greater than the VaR set by a given portfolio's management, to occur in a given day. Therefore, the 0.99-conditional quantile measures the maximum loss a portfolio can incur at 99% confidence level. It is clear from figure 4 that the 0.99-conditional quantile responds well to the distribution of the returns. We can also see from the figure that there are some few returns that exceed the VaR.

In figure 5, we have superimposed the $VaR_{t,0.99}$ and expected shortfall over the $VaR_{t,0.99}$ on the returns. The straight line represents the expected shortfall. We can observe from the figure that the straight line representing the expected shortfall is slightly above the line representing the $VaR_{t,0.99}$. Therefore, it is clear from the figure that VaR tells us the most we can expect to lose if a tail event does not occur while the expected shortfall tells us what we can expect to lose if a tail event does occur.

In our study, we have analyzed 963 MATATU returns. As per the definition of VaR at 99% confidence level, we expect, under normal market conditions, about 10 returns to exceed the $VaR_{t,0.99}$. In our results, we have obtained 9 exceedences and expected shortfall of 0.1688. Therefore, our risk measuring method is good.

5. Conclusion

In estimating the conditional mean and conditional volatility of the returns of our portfolio, we explored the exponential smoothing technique, whereby we assigned exponentially decreasing weights to the returns. We noted that exponential smoothing technique is easy to understand and apply. We proved that the estimators for the conditional mean and conditional volatility are consistent. We also proved that the estimators for the conditional mean and conditional volatility when conditional mean is known, are asymptotically normal. Further, we have given the estimators for the VaR and ES and proved that the VaR estimator is consistent.

Recommendations

Exponential smoothing techniques has a few demerits including that it lags i.e. the forecast will be behind as the trend increases over time and it also may fail to account for the dynamic changes at work in the real world, and the forecast may constantly require updating to respond to new information.

In this respect, we note that exponential smoothing technique may not be the best estimation technique and that other techniques, GARCH estimation may be explored.

We also suggest that the following work may be done in future as a continuation of this research:
a) The choice of the optimal smoothing parameter. In our case, we have just fixed its value.
b) The proof of the asymptotic normality for the conditional volatility estimator when the conditional mean is unknown. Also, the asymptotic normality for the VaR estimator should be shown.

References

[1] Aas, K., & Dimakos, X. K. (2004). Statistical modelling of financial time series: An introduction. Norwegian Computing Center; APPLIED RESEARCH AND DEVELOPMENT.

[2] Alexander, M. F. (1974). Third edition. In M. F. Alexander, Introduction to the Theory of Statistics (pp. 241-242, 295-296). McGraw-Hill.

[3] Amin, A. N. (2009). Reducing Emissions from Private Cars: Incentive measures for behavioural change*. United Nations Environment Programme.

[4] André Lucasa, X. Z. (2014). Score Driven Exponentially Weighted Moving Averages and Value-at-Risk Forecasting. University Amsterdam and Tinbergen Institute.

[5] Artzner, P. D. (1997). Thinking coherently. Risk.

[6] Artzner, P. F. (1997). "Thinking Coherently," Risk, 10 (11), Mathematical Finance, 68–71.

[7] Baki Billah, M. L. (2006, April–June). Exponential smoothing model selection for forecasting. International Journal of Forecasting, 22(2), 239-247.

[8] Boudoukh, J. R. (1997). Investigation of a class of volatility estimators. Journal of Derivatives, 4 Spring,, 63-71.

[9] Claudio H. da S. Barbedo, G. S. (2005). Evaluation of Foreign Exchange Risk Capital Requirement Models. Brazilian Review of Finance, Vol 3, No 2.

[10] Communications, M. o. (2004). Republic of Kenya. Nairobi.: Transformation of Road Transport Report.

[11] Dorothy Mccormick, W. M. (2012). Paratransit Operations And Regulation In Nairobi Matatu Business Strategies And The Regulatory Regime unpublished.

[12] Dorothy McCormick, W. M. (2013). Institutions and Business Strategies of Matatu Operators: A Case Study Report. Nairobi: African center for excellence forstudies in public and non-motorised Transport.

[13] Dr. Erik Winands, S. L. (2009). FORECASTING VOLATILITY IN THE STOCK MARKET. amsterdam: VU University.

[14] Gardner, E. J. (1985). Exponential smoothing: the state of the art. Journal of Forecasting, 1-28.

[15] Gizycki., J. A. (1999). Value at Risk: Volatility and Forecasting of the Variance Covariance Matrix.

[16] Huy-Nhiem Nguyen, Q. N. (2010). Exploring the Cost of Forecast Error in Inventory Systems. Fayetteville, AR 72701, U.S.A: Department of Industrial Engineering;University of Arkansas.

[17] Kibua1, P. O. (2004). EFFORTS TO IMPROVE ROAD SAFETY IN KENYA, ACHIEVEMENTS AND LIMITATIONS OF REFORMS IN THE MATATU INDUSTRY. Nairobi: Institute of Policy Analysis and Research (IPAR).

[18] Ladokhin, S. (2009). Volatility modeling in financial markets. Amsterdam: VU University.

[19] Legal Notices Nos. 161, 8. a. (2003 and 2004). Republic of Kenya.

[20] Madhavan, A., & Yang, J. (2003). Practical Risk Analysis for Portfolio Managers and Traders. ITG Inc.

[21] McCormick, D. W. (2011). Business Strategies of Matatu Operators in Nairobi: A Scoping Study ACET Project 14 Paratransit operations and regulation in Nairobi. University of Nairobi.

[22] McCormick, D. W. (2011). Institutions and business strategies of matatu operators in Nairobi: A scoping study ACET. African Centre of Excellence for Studies of Public and Non-Motorised Tr, Working Paper No. 14-2.

[23] MOA. (2014, november 20). matatuownersassociation. Retrieved from matatu owners association website: http://www.matatuownersassociation.com/index.php/about-us/our-profile

[24] Mwita, P. (2003). Semi-Parametric Estimation of Conditional Quantiles for Times Series with applications in finance. PhD Thesis, University of Kaiserslautern.

[25] Nyasetia, O. B. (2013). The influence of entreprenuarial personality, human capital and entry barriers on performance of entreprenuers in the informal transport business in nairobi, kenya. Unpublished.

[26] Pat. (2013, may 28). value-at-risk-with-exponential-smoothing. Retrieved from portfolio probe: http://www.portfolioprobe.com/2013/05/28/value-at-risk-with-exponential-smoothing/.

[27] Poon, S.-H. (2005). A Practical Guide to Forecasting Financial Market Volatility.pg1. John Wiley & Sons.

[28] Rau-Bredow, H. (2002). Value at Risk, Expected Shortfall, and Marginal Risk Contribution. Leo Weismantel Str. 4 D-97074 Würzburg.

[29] Roussas, G. G. (1973). A First Course in Mathematical Statistics. Addison-Wesley.

[30] S. Srinidhi, A. K. (2013). A conceptual model for demand forecasting and service positioning in the airline industry. Journal of Modelling in Management, 8(1), 123 – 139.

[31] S., E. (2015, March 3). Four Principles for Great sales forecast.. Retrieved from forbes: http: www.forbes.com

[32] Sirikhorn Klindokmai, P. N. (2014). Evaluation of forecasting models for air cargo. The International Journal of Logistics Management, 25(3), 635 - 655.

[33] Suwanvijit, W. C. (2009). Statistical model for short-term forecasting sparkling beverage sales in Southern Thailand. International Business and Economics Research Journal, 73-81.

[34] Tasche, C. A. (2001, may 9). Expected Shortfall: a natural coherent alternative to Value at Risk. Expected Shortfall: a natural coherent alternative to Value at Risk. Milano, Milano, Italy. Retrieved from http://arxiv.org/pdf/cond-mat/0105191.pdf

[35] Taylor, J. W. (2008). Using Exponentially Weighted Quantile Regression to Estimate Value at Risk and Expected Shortfall. Journal of Financial Econometrics, Vol. 6, pp.382-406.

[36] Tes´arov´a, B. V., & Gapko, P. P. (2012). Value at Risk: GARCH vs. Stochastic Volatility Models: Empirical Study. Prague: Charles University in Prague.

[37] Tian, G. J. (2008). Forecasting Volatility Using Long Memory and Comovements: An application to option valuation under SFAS 123R. Forthcoming Journal of Financial and Quantitative Analysis.

[38] Winands, S. L. (2009). Forecasting Volatility in the Stock Market. Amsterdam: VU University Amsterdam.

[39] Yamai, Y. Y. (2001). On the validity of value-at-risk: comparative analyses with Expected Shortfall. Institute for Monetary and Economic Studies, Bank of Japan.

[40] Yoshiba, Y. Y. (2002). On the Validity of Value-at-Risk: Comparative Analyses with Expected Shortfall. Research Division I, Institute for Monetary and Economic Studies, Bank of Japan.