Part I

Background

Chapter 2

Some Probability Essentials

2.1 Introduction[1]

In this chapter some of the essential background to probability theory is given. Although the background is basic, very few mathematicians, statisticians or probability theorists would be familiar with all the concepts presented.

In section 2.2 the basic concepts concerning a probability space is briefly stated. Moments are discussed in section 2.3.

Cumulative distribution functions and partial density functions are discussed in section 2.4. Some of the main theorems of this section is stated and proved. In section 2.5 the moments and other issues regarding the normal distribution is specified.

A short detour is taken in section 2.6 where returns series are discussed. The section ends with section 2.7 where some important hypothesis tests are discussed.

2.2 Probability Space

2.2.1 Probability Space

The triple $(\Omega, \mathcal{F}, P)$ is called a probability space. The set Ω is a non-empty set, $\mathcal{F} = \{\mathcal{F}_t\}_{t \in N, t \geq 0}$ is filtration of $\sigma-$algebras $\mathcal{F}_t$ defined on Ω and P is a probability measure on $\mathcal{F}$.

A function $Z_t : \Omega \to \mathbb{R}$, on the probability space, is called a stochastic process.

2.2.2 $\sigma-$algebra

A family of subsets $\mathcal{F}$ of a set Ω is called a $\sigma-$algebra if the following holds:

[1]For further discussions on probability theory and measure theoretical aspects see [31], [7], [3] and [27]. [4], [13], [17], [26] and [32] are also useful.

1. $\emptyset \in \mathcal{F}$
2. If $X \in \mathcal{F}$ then $\Omega \backslash X \in \mathcal{F}$
3. If (X_n) is a sequence of sets in Ω then $\bigcup_{n=1}^{\infty} X_n \in \mathcal{F}$.

2.2.3 Borel Sets in $\mathbb{R}$

The Borel sets (one thing) is the smallest $\sigma-$algebra generated by all the open sets in $\mathbb{R}$.

2.2.4 Filtration

$\mathcal{F} = \{\mathcal{F}_t\}_{t \in N, t \geq 0}$ is a filtration of $\sigma-$algebras with the following properties

1. $\mathcal{F}_0$ contains all null sets
2. $\mathcal{F}_t = \cap_{s:s>t} \mathcal{F}_s$ for $t \geq 0$

Property 2 is called right continuous. A filtration with these properties is said to satisfy the "usual conditions".

2.2.5 Measurability and Adaptedness

Definition 2.2.1 *A stochastic process, $Z_t : \Omega \rightarrow \mathbb{R}$, is said to be measurable with respect to a σ-algebra $\mathcal{F}$ if*

$$\{Z_t \in B\} \in \mathcal{F}$$

for every Borel set $B \in B(\mathbb{R})$.

Definition 2.2.2 *A stochastic process, Z_t, is said to be adapted to a filtration $\mathcal{F}$ if Z_t is $\mathcal{F}_t$ measurable for all $t \in \mathbb{R}$.*

Remark 2.2.3 *Throughout this dissertation only real-valued stochastic processes defined on $(\Omega, \mathcal{F}, P)$ will be considered, that is $X : \Omega \rightarrow \mathbb{R}$.*

Remark 2.2.4 *A stochastic process at a specific time is often referred to as a random variable.*

Remark 2.2.5 *Take note that a stochastic process/random variable is defined in terms of a probability space.*

2.2.6 Almost everywhere

Definition 2.2.6 *Two functions, f and g are equal almost everywhere (sometimes called almost surely) if*

$$f(x) = g(x)$$

for all $x \notin N \in \mathcal{F}$ where $P(N) = 0$. Almost everywhere is abbreviated by a.e.

Definition 2.2.7 *A sequence of functions (f_n) converges to f almost everywhere if there exists a set $N \in \mathcal{F}$ with $P(N) = 0$ such that $f(x) = \lim f_n(x)$ for all $x \notin N$.*

2.3 Moments and Stationarity

2.3.1 Expected Value

Definition 2.3.1 *A random variable $X \in \Omega \to \mathbb{R}$ is said to be integrable if*

$$\int_\Omega |X| \, dP < \infty$$

The family of integrable random variables are denoted by $L^1(\Omega, \mathcal{F}, P)$ or in this dissertation L^1 for short.

Definition 2.3.2 *For any $X \in L^1(\Omega, \mathcal{F}, P)$,*

$$E[X] := \int_\Omega X dP$$

is called the expected value of X.

Remark 2.3.3 *The expected value of a random variable from a symmetric distribution is often called the mean or average.*

Remark 2.3.4 *For a probability space with density function f and integrable Borel function $h : \mathbb{R} \to \mathbb{R}$,*

$$E[h(X)] = \int_\mathbb{R} h dP_X = \int_{-\infty}^{\infty} h(x) f(x) \, dx$$

Proof. See Brzezniak et al. [7]. ∎

2.3.2 Conditional Expectation

We can call the filtration $\mathcal{F}_t \subset \mathcal{F}$, the σ–algebra that contains all the information available to an investor at time t. A σ-algebra can also be a condition in a conditional expectation.

Definition 2.3.5 $(\Omega, \mathcal{F}, P)$. *The conditional expectation given a σ–algebra is: for an integrable random variable X and σ–algebra $\Phi \subseteq \mathcal{F}$*

$$E[X \mid \Phi]$$

where $E[X \mid \Phi] : \Omega \to \mathbb{R}$ is unique P a.s. satisfying

1. *$E[X \mid \Phi]$ is Φ measurable*
2. *$\int_H E[X \mid \Phi]\, dP = \int_H X dP$, for all $H \in \Phi$.*

Theorem 2.3.6 *For X, $Y \in L^1$, $a, b \in \mathbb{R}$ and σ – algebra $\Phi \subset \mathcal{F}$ the conditional expectation has the following basic properties (all equalities a.s.):*

1. *$E[aX + bY \mid \Phi] = aE[X \mid \Phi] + bE[Y \mid \Phi]$*
2. *$E[E[X \mid \Phi] \mid \Psi] = E[X \mid \Psi]$ where $\Phi \subset \Psi$ is also a σ – algebra*
3. *$E[X \mid \Phi] = X$ if X is Φ measurable*
4. *$E[X \mid \Phi] = E[X]$ if X is independent of Φ*

Proof. See Appendix B in Oksendal [27]. ■

2.3.3 Variance, Conditional Variance and Standard Deviation

Definition 2.3.7 $(\Omega, \mathcal{F}, P)$. *A random variable $X \in \Omega \to \mathbb{R}$ is said to be square integrable if*

$$\int_\Omega X^2 dP < \infty$$

The family of square integrable random variables are denoted by $L^2(\Omega, \mathcal{F}, P)$ or L^2 for short.

Definition 2.3.8 $(\Omega, \mathcal{F}, P)$. *The variance of a square integrable random variable X is defined as*

$$\begin{aligned} Var[X] &= E\left[(X - E[X])^2\right] \\ &= E[X^2] - (E[X])^2 \\ &= \int_H X^2 dP + \left(\int_H X dP\right)^2 > 0 \end{aligned}$$

Definition 2.3.9 $(\Omega, \mathcal{F}, P)$. *The standard deviation of a square integrable random variable X is defined as*

$$Std[X] = \sqrt{Var[X]}$$

The conditional variance and its properties follows directly from that of the conditional expected value:

Definition 2.3.10 $(\Omega, \mathcal{F}, P)$. *For any square integrable random variable X, the conditional expected value of X given a $\sigma - algebra$ $H \subseteq \mathcal{F}$, is*

$$\begin{aligned} Var[X \mid H] &= E\left[(X - E[X \mid H])^2 \mid H\right] \\ &= E\left[X^2 - 2XE[X \mid H] + (E[X \mid H])^2 \mid H\right] \text{ a.s. } P \\ &= E\left[X^2 \mid H\right] - (E[X \mid H])^2 \text{ a.s. } P \end{aligned}$$

Theorem 2.3.11 $(\Omega, \mathcal{F}, P)$. *For a square integrable random variable Z and $a, c \in \mathbb{R}$, the conditional variance of $K = aX + c$, given a $\sigma - algebra$ $H \subseteq \mathcal{F}$, is*

$$Var(K \mid H) = a^2 Var(X \mid H) \text{ a.s. } P$$

Proof. *Equalities almost surely*

$$\begin{aligned} & Var[K \mid H] \\ =\ & E\left[(K)^2 \mid H\right] - (E[K \mid H])^2 \\ =\ & E[a^2X^2 + 2acX + c^2 \mid H] - (E[aX \mid H] + c)^2 \\ =\ & a^2E[X^2 \mid H] + 2acE[X \mid H] + c^2 - a^2(E[X \mid H])^2 \\ & -2ac(E[X \mid H]) - c^2 \\ =\ & a^2E[X^2 \mid H] - a^2(E[X \mid H])^2 \\ =\ & a^2 Var(X \mid H). \end{aligned}$$

■

Theorem 2.3.12 $(\Omega, \mathcal{F}, P)$.*For two square integrable random variable X and Y and $a, b, c \in \mathbb{R}$, the conditional variance of $Z = aX + bY + c$*

$$Var(Z \mid H) = a^2 Var(X \mid H) + b^2 Var(Y \mid H) + 2abCov(X, Y \mid H) \text{ a.s. } P$$

Proof. *From theorem 2.3.11. Equalities almost surely P*

$$
\begin{aligned}
& Var\left[aX+bY+c \mid H\right] \\
= \;& Var\left[aX+bY \mid H\right] \\
= \;& E\left[(aX+bY)^2 \mid H\right] - \left(E\left[aX+bY \mid H\right]\right)^2 \\
= \;& a^2E\left[X^2 \mid H\right] + 2abE\left[XY \mid H\right] + b^2E\left[Y^2 \mid H\right] \\
& -\left(aE\left[X \mid H\right] + bE\left[Y \mid H\right]\right)^2 \\
= \;& a^2E\left[X^2 \mid H\right] + 2abE\left[XY \mid H\right] + b^2E\left[Y^2 \mid H\right] - a^2\left(E\left[X \mid H\right]\right)^2 \\
& -2abE\left[X \mid H\right]E\left[Y \mid H\right] - b^2\left(E\left[Y \mid H\right]\right)^2 \\
= \;& a^2Var\left(X \mid H\right) + b^2Var\left(Y \mid H\right) \\
& +2ab\left(E\left[XY \mid H\right] - E\left[X \mid H\right]E\left[Y \mid H\right]\right)
\end{aligned}
$$

In the next section covariances will be properly defined, for now assume

$$Cov\left(X,Y \mid H\right) = E\left[XY \mid H\right] - E\left[X \mid H\right]E\left[Y \mid H\right].$$

Thus

$$
\begin{aligned}
& Var\left[aX+bY+c \mid H\right] \\
= \;& a^2Var\left(X \mid H\right) + b^2Var\left(Y \mid H\right) + 2abCov\left(X,Y \mid H\right)
\end{aligned}
$$

■

2.3.4 Covariance and Autocovariance

Definition 2.3.13 $(\Omega, \mathcal{F}, P)$. *For two square integrable random variables X and Y in our probability space, the covariance between X and Y is*

$$
\begin{aligned}
cov\left[X,Y\right] &= E\left[\left(X-E\left[X\right]\right)\left(Y-E\left[Y\right]\right)\right] \\
&= E\left[XY\right] - E\left[X\right]E\left[Y\right]
\end{aligned}
$$

Definition 2.3.14 $(\Omega, \mathcal{F}, P)$. *For a square integrable stochastic process $(X_t)_{t \in N}$, adapted to $\mathcal{F}$, the covariance between X_t and X_{t-k} for any $t, k \in N$ is*

$$
\begin{aligned}
cov\left[X_t, X_{t-k}\right] &= E\left[\left(X_t - E\left[X_t\right]\right)\left(X_{t-k} - E\left[X_{t-k}\right]\right)\right] \\
&= E\left[X_tX_{t-k}\right] - E\left[X_t\right]E\left[X_{t-k}\right]
\end{aligned}
$$

The covariance between elements of the same stochastic process is called the autocovariance.

The conditional covariance and autocovariance can be defined in a similar fashion as the conditional variance, bearing in mind that conditional covariances are random variables.

2.3.5 Correlation and Autocorrelation

Definition 2.3.15 $(\Omega, \mathcal{F}, P)$. *For two square integrable random variables X and Y the correlation between X and Y is*

$$cor[X,Y] = \frac{cov[X,Y]}{\sqrt{var[X]\,var[Y]}}$$

Definition 2.3.16 $(\Omega, \mathcal{F}, P)$. *For a square integrable stochastic process $(X_t)_{t \in N}$, adapted to $\mathcal{F}$, the correlation between X_t and X_{t-k} for any $t, k \in N$ is*

$$cor[X_t, X_{t-k}] = \frac{cov[X_t, X_{t-k}]}{\sqrt{var[X_t]\,var[X_{t-k}]}}$$

The correlation between elements of the same stochastic process is called the autocorrelation.

2.3.6 Lag

Definition 2.3.17 *Consider a stochastic process, say $(X_t)_{t \in N}$. At any time step t a lag of size k is an integer that represents the process at time $t - k$, X_{t-k}.*

2.3.7 Higher Moments

Definition 2.3.18 $(\Omega, \mathcal{F}, P)$. *The r^{th} moment of a random variable X (about its mean) is*

$$E\left[(X - E[X])^r\right]$$

The first moment of a random variable is defined as its mean. The second moment of a random variable is its variance. The second moments of a stochastic process also include the autocovariances. The third moment of a random variable is skewness and the fourth is kurtotsis. For a stochastic process $(X_t)_{t \in N}$ the set of r^{th} moments can be defined as

$$\left\{ E\left[\prod_{i=1}^{r} (X_{k_i} - E[X_{k_i}])\right] \mid \text{for all } k_i \in N \right\}$$

2.3.8 Stationarity

Definition 2.3.19 *A stochastic process is called stationary if all of its moments are constants.*

Definition 2.3.20 *A stochastic process is called weakly stationary if its first and second moments are constant. This means that its mean is constant and for every lag k and time t the $cov[X_t, X_{t-k}]$ is a constant.*

2.4 Cumulative Distribution Function and Probability Density Function

Definition 2.4.1 *The (cumulative) distribution function (cdf) of a random variable* $X : \Omega \to \mathbb{R}$ *is defined as*

$$F(x) = P\{X \leq x\}$$

Theorem 2.4.2 *The cdf* F *of a random variable* $X : \Omega \to \mathbb{R}$ *has the following properties*

1. $0 \leq F \leq 1$
2. $\lim_{x \to -\infty} F(x) = 0$ *and* $\lim_{x \to \infty} F(x) = 1$
3. F *is right-continuous,* $F(x) = \lim_{x_n \to x} F(x_n)$ *for a decreasing sequence* x_n
4. F *is increasing.*

Proof. *See Brzezniak et al. [7].* ■

Theorem 2.4.3 *If* $g : \mathbb{R} \to \mathbb{R}$ *is integrable then*

$$E[g(x)] = \int_{-\infty}^{\infty} g(x)\, dF(x)$$

Proof. *A result of exercise 1.7 in Brzezniak et al. [7].* ■

Theorem 2.4.4 *A measurable function* $f(x) \geq 0$ *on* $\mathbb{R}$ *such that for any Borel measurable set* B

$$P\{X \in B\} = \int_B f(y)\, dy$$

is called the (probability) density function (pdf) of X. *The pdf can in particular also be written in terms of the cdf of* X, F

$$F(x) = \int_{-\infty}^{y} f(y)\, dy$$

Proof. *See Brzezniak et al. [7].* ■

Theorem 2.4.5 *If* X *has a continuous pdf* f *then*

$$\frac{d}{dx} F(x) = f(x)$$

Proof. *A result from the fundamental theorem of calculus.* ■

2.4.1 Joint Continuous Distributions

The joint distribution of a k-dimensional random variable,

$$\mathbf{X} = (X_1, X_2, ..., X_k),$$

is a measure $P_{\mathbf{X}}$ on $\mathbb{R}^n$ such that for any Borel set, $\mathbf{B} \in \mathbb{R}^n$

$$P_{\mathbf{X}}(\mathbf{B}) = P\{X \in \mathbf{B}\}$$

If the random variables of X are independently distributed then

$$P_{\mathbf{X}}(\mathbf{B}) = \prod_{i=1}^{k} P\{X_i \in B_i\}$$

where

$$\mathbf{B} = \begin{bmatrix} B_1 \\ ... \\ B_k \end{bmatrix}.$$

Definition 2.4.6 *The joint probability density function (joint pdf) of a k-dimensional random variable,*

$$\mathbf{X} = (X_1, X_2, ..., X_k),$$

is a Borel function

$$f(x_1, x_2, ..., x_k) : \mathbb{R}^n \to \mathbb{R}$$

such that

$$P_{\mathbf{X}}(\mathbf{B}) = \int_{\mathbf{B}} f(t_1, t_2, ..., t_k)\, dt_1 ... dt_k \tag{2.1}$$

Definition 2.4.7 *The joint cumulative distribution function (joint cdf) of a k-dimensional random variable,*

$$\mathbf{X} = (X_1, X_2, ..., X_k),$$

is

$$F(x_1, ..., x_k) = P[X_1 \leq x_1, ..., X_k \leq x_k]$$

If the random variables of $\mathbf{X}$ are independently distributed then

$$f(x_1, ..., x_k) = f(x_1) ... f(x_k)$$

and

$$F(x_1, ..., x_k) = F(x_1) ... F(x_k)$$

which follows directly from the case of independence of $P_{\mathbf{X}}$.

Theorem 2.4.8 *The joint cdf of a k-dimensional random variable* $\mathbf{X}$ *can be written in terms of the joint pdf of as follows*

$$F(x_1, x_2, ..., x_k) = \int_{-\infty}^{x_k} ... \int_{-\infty}^{x_1} f(t_1, t_2, ..., t_k)\, dt_1 ... dt_k \tag{2.2}$$

Proof. From definition 2.4.7 and the fact that $(-\infty, x_i]$ is a Borel set for every applicable i it is clear that the joint cdf of X is a special case of the joint probability of X. Equation 2.2 follows directly from 2.1. ■

Theorem 2.4.9 *If* $X, Y \in \mathbb{R}$ *are independent random variables and* $g(x)$ *and* $h(y)$ *are functions then*

$$E[g(X)h(Y)] = E[g(X)]E[h(Y)]$$

and

$$cov[g(X), h(Y)] = 0$$

Proof. With a joint pdf $f(x, y)$

$$\begin{aligned} E[g(X)h(Y)] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x)h(y)f(x,y)\,dxdy \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x)h(y)f_1(x)f_2(y)\,dxdy \end{aligned}$$

due to independence. The cdfs of X and Y are f_1 and f_2 respectively, then

$$\begin{aligned} E[g(X)h(Y)] &= \int_{-\infty}^{\infty} h(y)f_2(y)\int_{-\infty}^{\infty} g(x)f_1(x)\,dxdy \\ &= \int_{-\infty}^{\infty} g(x)f_1(x)\,dx\int_{-\infty}^{\infty} h(y)f_2(y)\,dy \\ &= E[g(X)]E[h(Y)] \end{aligned}$$

The covariance can be expressed as

$$\begin{aligned} & cov[g(X), h(Y)] \\ =\ & E[g(X)h(Y)] - E[g(X)]E[h(Y)] \\ =\ & 0 \end{aligned}$$

■

2.5 The Normal Distribution and its Moment Generating Function

2.5.1 The Normal Distribution

The normal distribution, the most frequently used statistical distribution, was first published by Abraham de Moivre (1733).

A normal random variable $X \in \mathbb{R}$, with mean $\mu \in \mathbb{R}$ and variance $\sigma^2 \in \mathbb{R}^+$ is denoted by

$$X \sim N\left(\mu, \sigma^2\right)$$

The probability density function (pdf) of the normal distribution is

$$f\left(x; \mu, \sigma^2\right) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$$

Another way to define the pdf in terms of the probability space $(\Omega, \mathcal{F}, P)$ is as follows

$$P(A) = \int_A dP = \int_A f\left(x; \mu, \sigma^2\right) dx$$

The cumulative distribution function (cdf) of the normal distribution is given by

$$\begin{aligned} F\left(z; \mu, \sigma^2\right) &= P\{X \leq z\} \\ &= \int_{\{X \leq z\}} \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] dx \\ &= \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] dx \end{aligned}$$

The standard normal distribution, frequently used in this dissertation is defined as the normal distribution with zero mean and a variance of one, $N(0,1)$. The standard normal distribution's pdf is

$$f(x; 0, 1) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$$

and cdf is

$$F(x; 0, 1) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx.$$

2.5.2 Moments of the Normal Distribution

Consider a normally distributed random variable $X \sim N\left(\mu, \sigma^2\right)$ with probability density function

$$f\left(x; \mu, \sigma^2\right) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$$

then the random variable

$$Y = \frac{X-\mu}{\sigma} \sim N(0,1)$$

Definition 2.5.1 *The moment generating function of X is defined as*

$$M_X(t) = E\left(e^{tX}\right)$$

Theorem 2.5.2 *The moment generating function of $X \sim N\left(\mu, \sigma^2\right)$ is*

$$M_X(t) = \exp\left(\mu t + \frac{(\sigma t)^2}{2}\right)$$

Proof.

$$\begin{aligned}
M_X(t) &= E\left(e^{tX}\right) \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{tx} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} \exp\left[tx - \frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] dx \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(\frac{x^2 - 2\left(\mu + \sigma^2 t\right)x + \mu^2}{-2\sigma^2}\right) dx \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(\frac{\left(x - \mu + \sigma^2 t\right)^2 - 2\mu\sigma^2 t - \left(\sigma^2 t\right)^2}{-2\sigma^2}\right) dx \\
&= \exp\left(\mu t + \frac{(\sigma t)^2}{2}\right) \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(\frac{\left(x - \mu + \sigma^2 t\right)^2}{-2\sigma^2}\right) dx \\
&= \exp\left(\mu t + \frac{(\sigma t)^2}{2}\right)
\end{aligned}$$

■

Theorem 2.5.3 *Moments about the mean of $X \sim N\left(\mu, \sigma^2\right)$. If r is even then*

$$E\left[(X-\mu)^r\right] = \frac{(2r)!\sigma^{2r}}{r!2^r},$$

if r is odd then

$$E\left[(X-\mu)^r\right] = 0$$

Proof. The

$$
\begin{aligned}
M_{X-\mu} &= \exp\left(\frac{(\sigma t)^2}{2}\right) \\
&= \sum_{n=0}^{\infty} \frac{\left(\frac{(\sigma t)^2}{2}\right)^n}{n!} \\
&= \sum_{n=0}^{\infty} \frac{\sigma^{2n} t^{2n}}{2^n n!} \\
&= \sum_{n=0}^{\infty} \frac{\sigma^{2n} t^{2n} (2n!)}{2^n n! (2n!)} \\
&= \sum_{n=0}^{\infty} \frac{\sigma^{2n} (2n!)}{2^n n!} \frac{t^{2n}}{(2n!)}
\end{aligned}
$$

The second line is due to the Maclaurin series expansion for e. Note that only positive integers are contained in the last line Theorem 2.5.1 in [2] states that

$$M_V(t) = 1 + \sum_{n=1}^{\infty} E[V^n] \frac{t^n}{n!}$$

for a random variable V. Thus

$$E[(X-\mu)^r] = 0$$

if r is odd and

$$E[(X-\mu)^r] = \frac{\sigma^r r!}{2^{r/2}(r/2)!}$$

if r is even. ■

The following characteristics of random variable $X \sim N(\mu, \sigma^2)$ follows from theorem 2.5.3:

1. The skewness of X is

$$E\left[(X-\mu)^3\right] = 0$$

2. The kurtosis of X is

$$
\begin{aligned}
E\left[(X-\mu)^4\right] &= \frac{\sigma^4 4!}{2^2 (2)!} \\
&= 3\sigma^4
\end{aligned}
$$

and thus if $\sigma^2 = 1$

$$E\left[(X-\mu)^4\right] = 3$$

2.5.3 Chi-square Distribution

Definition 2.5.4 *If random variable Y is chi-square distributed with v degrees of freedom then*

$$Y \sim \chi^2(v)$$

where the chi-square distributed is a special case of the gamma distribution

$$\chi^2(v) \sim GAM\left(2, \frac{v}{2}\right)$$

Theorem 2.5.5 *A random variable $Y \sim \chi^2(v)$ has the following characteristics*

1. *Probability density function*

$$f(y) = \frac{1}{2^{v/2}\Gamma(v/2)} y^{v/2-1} e^{-x/2},$$

where Γ is the gamma function

$$\Gamma(\kappa) = \int_0^\infty t^{k-1} e^{-t} dt$$

for all $\kappa > 0$.

2. *Moment generating function*

$$M_Y(t) = (1-2t)^{-v/2}$$

3. *Moments about the mean*

$$E[Y^r] = 2^r \frac{\Gamma(v/2+r)}{\Gamma(v/2)}$$

4. *Expected value*

$$E[Y] = v$$

5. *Variance*

$$Var[Y] = 2v$$

Proof. *Results follow from the gamma distribution. See Bain [2]* ■

Theorem 2.5.6 *If*

$$X \sim N\left(\mu, \sigma^2\right)$$

then

$$Z^2 = \left(\frac{X-\mu}{\sigma}\right)^2 \sim \chi^2(1)$$

Proof. *The moment generating function of* Z^2

$$\begin{aligned} M_{Z^2} &= E\left[e^{tZ^2}\right] \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(tz^2\right) \exp\left(-\frac{1}{2}z^2\right) dz \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(tz^2 - \frac{1}{2}z^2\right) dz \\ &= \frac{1}{\sqrt{1-2t}} \int_{-\infty}^{\infty} \frac{\sqrt{1-2t}}{\sqrt{2\pi}} \exp\left(\frac{z^2(1-2t)}{2}\right) dz \\ &= \frac{1}{\sqrt{1-2t}} \end{aligned}$$

which is the moment generating function of the chi-square distribution. ∎

Definition 2.5.7 *If* $Z^2 \sim \chi^2(1)$ *then*

$$(Z-\lambda)^2$$

is noncentral chi-square distributed with 1 *degree of freedom and non-centrality parameter* λ.

Theorem 2.5.8 *The expected value of a noncentral chi-square distributed random variable is*

$$E\left[(Z-\lambda)^2\right] = 1 + \lambda^2$$

where $Z^2 \sim \chi^2(1)$.

Proof.

$$\begin{aligned} E\left[(Z-\lambda)^2\right] &= E\left[Z^2 - 2\lambda Z + \lambda^2\right] \\ &= E\left[Z^2\right] - 2\lambda E[Z] + \lambda^2 \\ &= 1 + \lambda^2 \end{aligned}$$

since $E[Z] = 0$ because $Z \sim N(0,1)$. ∎

i 16542502
b15954341

Theorem 2.5.9 *The expected value of a noncentral chi-square distributed random variable is*

$$E\left[(Z-\lambda)^4\right] = 3 + 6\lambda^2 + \lambda^4$$

where $Z^2 \sim \chi^2(1)$.

Proof.

$$\begin{aligned} E\left[(Z-\lambda)^4\right] &= E\left[\left(Z^2 - 2\lambda Z + \lambda^2\right)^2\right] \\ &= E\left[Z^4 - 4Z^3\lambda + 6Z^2\lambda^2 - 4\lambda^3 Z + \lambda^4\right] \\ &= E\left[Z^4\right] - 4\lambda E\left[Z^3\right] + 6\lambda^2 E\left[Z^2\right] - 4\lambda^3 E[Z] + \lambda^4 \\ &= 3 + 6\lambda^2 + \lambda^4 \end{aligned}$$

This is done by remembering that

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

thus the expected value of Z is

$$E[Z] = 0$$

the skewness is

$$E\left[Z^3\right] = 0$$

and the kurtosis is

$$E\left[Z^4\right] = 3$$

■

2.6 The Return Series and Lognormal Distribution

2.6.1 Returns Series

The financial value of a company or fund is represented by its (stock) price. The stock price has a clear, time dependent trend. It is hard to model series with trends, at least in an objective, scientific sense. To remove this trend, the financial time series is transformed into a series with "manageable" mean, a returns series. This is done with difference equations.

It will be proved that the returns series still has the same variance as the original series. The returns series is of great importance in risk management and derivatives pricing.

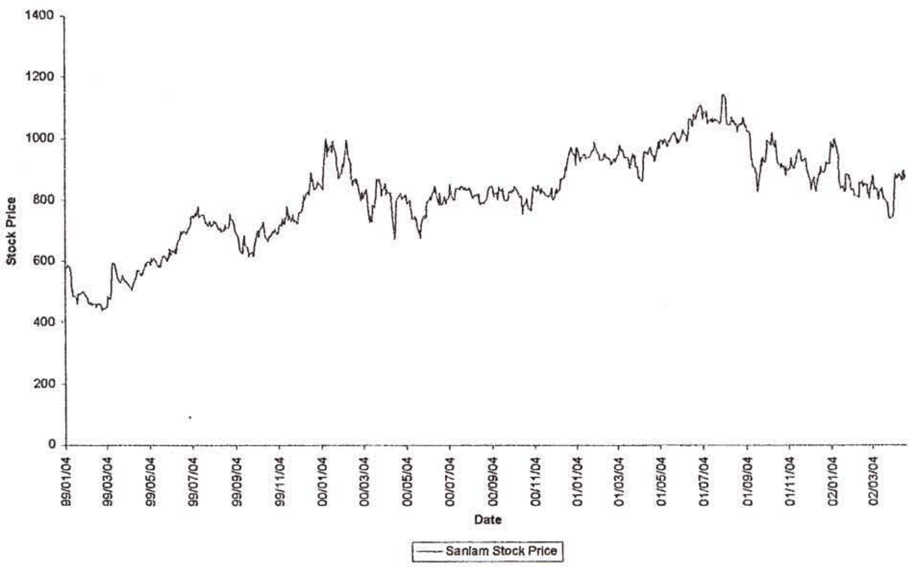

Figure 2.1: The stock price of Sanlam from 1999/01/05 to 2002/04/19.

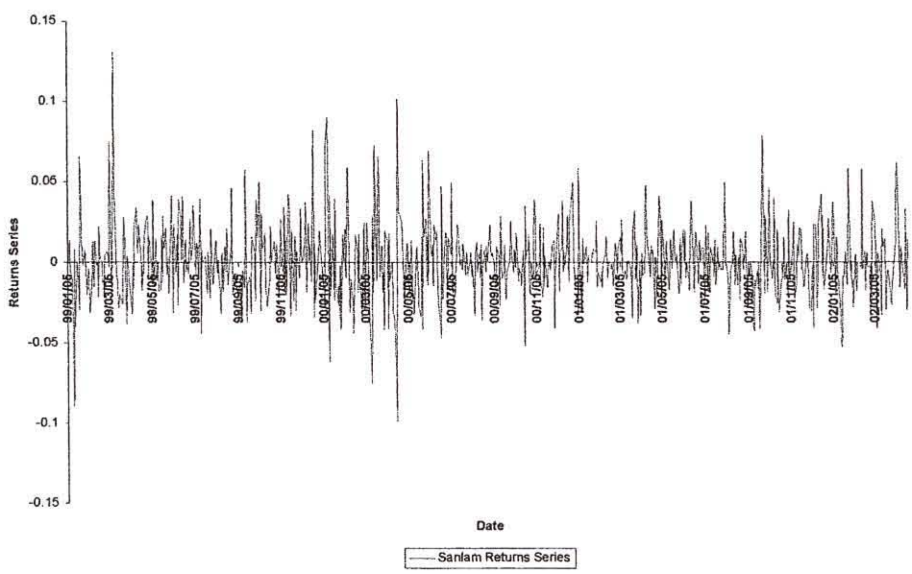

Figure 2.2: The returns series of Sanlam from 1999/01/05 to 2002/04/19.

2.6.2 The Arithmetic Returns Series

Definition 2.6.1 *The arithmetic returns series, for process* (S_t) *is defined as*

$$r_t = \frac{S_t - S_{t-1}}{S_{t-1}}$$

2.6.3 The Geometric Returns Series

The geometric returns series, for process (S_t) is defined as

$$\begin{aligned} r_t &= lnS_t - lnS_{t-1} \\ &= ln\left(\frac{S_t}{S_{t-1}}\right) \end{aligned}$$

The relationship between the geometric and arithmetic series, by the Taylor series expansion, are as follows:

$$\begin{aligned} ln\left(\frac{S_t}{S_{t-1}}\right) &= ln\left(\frac{S_t}{S_{t-1}} - 1 + 1\right) \\ &= ln\left(\frac{S_t - S_{t-1}}{S_{t-1}} + 1\right) \\ &\approx \frac{S_t - S_{t-1}}{S_{t-1}} \end{aligned}$$

if $\left|\frac{S_t - S_{t-1}}{S_{t-1}}\right| < 1$.

The geometric returns series will be considered in this dissertation.

Theorem 2.6.2 *If we assume that a returns series is normally distributed then the log of the stock process is also normal, and vice versa.*

Proof.

$$\begin{aligned} ln\left(\frac{S_1}{S_0}\right) + ln\left(\frac{S_2}{S_1}\right) + ln\left(\frac{S_2}{S_1}\right) + ... + ln\left(\frac{S_t}{S_{t-1}}\right) &= ln\left(\frac{S_t}{S_0}\right) \\ &= lnS_t - lnS_0 \end{aligned}$$

The sum of normally distributed random variables are also normal and we assume that S_0 is known. ■

2.6.4 Lognormal Distribution

$(\Omega, \mathcal{F}, P)$. A random variable $X \in \mathbb{R}$, with mean μ and variance σ^2 is said to be lognormally distributed if $\ln(X)$ is normally distributed.

It's often observed that stock prices are lognormally distributed. In chapter 5.4.1 we deduce, given the assumed process 5.10, that a stock price

S_t can be defined in terms of an initial stock price S_0 and Brownian motion W_t ,

$$S_t = S_0 \exp\left(\left(\mu - \frac{1}{2}\sigma^2\right)t + \sigma W_t\right) \tag{2.3}$$

Taking the logarithm on both sides of equation 2.3 yields

$$\ln S_t \sim N\left(\ln S_0 + \left(\mu - \frac{1}{2}\sigma^2\right)t, \sigma^2 t\right)$$

The return on stock S_t is defined as $\ln\left(\frac{S_t}{S_{t-1}}\right)$ which clearly has the distribution

$$\ln \frac{S_t}{S_{t-1}} \sim N\left(\mu - \frac{1}{2}\sigma^2, \sigma^2\right)$$

It is possible to test with the Jarque-Bera test for normality whether the return is in reality normally distributed.

2.7 Hypothesis Testing[2]

Hypothesis tests are done to verify whether the properties of an observed series, say $\{\hat{\varepsilon}_t\}_{t\in N}$, are consistent with assumed properties under a model The properties that need to be tested include tests for normality, autocorrelation and heteroscedasticity.

The formal procedure for conducting a hypothesis test involves a statement of the null hypothesis and an alternative hypothesis. The sample estimate on which the decision to reject or not reject the null hypothesis comes from the sample space. The Neyman-Pearson methodology [20] involves partitioning the sample space into two regions. If the sample estimate falls in the critical region, the null hypothesis is rejected. If it falls in the acceptance region, it's not.

2.7.1 Jarque-Bera Test for Normality

The Jarque-Bera tests whether observations are not likely to have come from the normal distribution.

Define for n observations the following

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{t=1}^{n} \hat{\varepsilon}_t^2, \tag{2.4}$$

$$\hat{\mu}_3 = \frac{1}{n}\sum_{t=1}^{n} \hat{\varepsilon}_t^3, \tag{2.5}$$

[2] Suggested reading: [1], [2], [18] and [24].

$$\hat{\mu}_4 = \frac{1}{n}\sum_{t=1}^{n}\hat{\varepsilon}_t^4 \tag{2.6}$$

In equations 2.4 to 2.6 are the second, third and fourth moments of $\hat{\varepsilon}_t$ respectively.

The skewness is defined as

$$skewness = s = \frac{\hat{\mu}_3}{\hat{\sigma}^2}$$

and the kurtosis as

$$kurtosis = k = \frac{\hat{\mu}_4}{\hat{\sigma}^2}$$

The Jarque-Bera statistic is defined as

$$\begin{aligned} JB &= n\left(\frac{\hat{\mu}_3^2}{6\hat{\sigma}^6} + \frac{\left(\hat{\mu}_4 - 3\hat{\sigma}^4\right)^2}{24\hat{\sigma}^8}\right) \\ &= \frac{n}{6}\left(s^2 + 2\left(k-3\right)^2\right) \end{aligned}$$

$$JB \sim \chi^2\left(2\right)$$

The null hypothesis is

$$H_0 : s = 0 \; and \; k = 3$$

against the alternative

$$H_1 : reject \, H_0$$

Remark 2.7.1 *Many text books and computer packages calculates the adjusted kurtosis, that is the* $k - 3$.

2.7.2 Autocorrelation

Durbin-Watson

The most famous test for autocorrelation is the Durbin-Watson test

$$\frac{\sum_{r=2}^{t}\left(\hat{\varepsilon}_r - \hat{\varepsilon}_{r-1}\right)^2}{\sum_{r=1}^{t}\hat{\varepsilon}_r^2}$$

No exact distribution for this test is available.

Ljung-Box

For a series with m observations the Ljung-Box statistic over K lags is

$$m(m+2)\sum_{k=1}^{K}\frac{\tilde{\rho}_k^2}{m-k}\sim \varkappa^2(K)$$

where $\tilde{\rho}_k^2$ is the observed autocorrelation at lag k given by

$$\tilde{\rho}_k=\frac{\sum_{r=k+1}^{t}\left(\hat{\varepsilon}_t^2-\tilde{\sigma}^2\right)\left(\hat{\varepsilon}_{t-k}^2-\tilde{\sigma}^2\right)}{\sum_{r=1}^{t}\left(\hat{\varepsilon}_t^2-\tilde{\sigma}^2\right)^2}$$

where $\hat{\varepsilon}_t$ is the observed return at time t and $\tilde{\sigma}^2$ is the sample variance.

2.7.3 Volatility Clustering

Many financial time series and also the Black-Scholes option pricing model make the assumption of constant volatility. Empirical evidence indicates that volatility of financial instruments tends to be dynamic. Volatility levels tend to alternate between periods of higher volatility and more tranquil periods. This clustering together of volatility levels for a period of time is called volatility clustering. Volatility clustering is due to the strong autocorrelation of squared returns or absolute returns. The Box-Pierce Lagrange multiplier test for the significance of first-order autocorrelation in squared returns, $\hat{\varepsilon}_t^2$, is

$$\frac{\sum_{t=2}^{T}\hat{\varepsilon}_t^2\hat{\varepsilon}_{t-1}^2}{\sum_{t=2}^{T}\hat{\varepsilon}_t^4}$$

The Lagrange multiplier tests are chi-squared distributed with T degrees of freedom.

2.7.4 The Leverage Effect

Volatility tends to be higher in a falling market, than in a rising market. Similarly volatility tends to be higher after a large negative return than after a large positive return, for an individual stock. The reason for this is that when a stock price falls, the leverage or debt/equity ratio increases. In laymen's terms, the part of the company's assets "owned" by the creditors increases, leaving less for the shareholders. This causes more uncertainty in the stock price.

An asymmetric version of the Lagrange multiplier test is used to investigate the influence of the leverage effect, and asymmetric returns levels in general

$$\frac{\sum_{t=2}^{T}\hat{\varepsilon}_t^2\hat{\varepsilon}_{t-1}}{\sum_{t=2}^{T}\hat{\varepsilon}_t^2\hat{\varepsilon}_{t-1}^2}$$

where $\hat{\varepsilon}_t$ is the observed return at time t.