Chapter 1

Introduction

A typical goal in the financial arena is to attempt to predict the future based upon knowledge of the past, in order to make profitable trades. In this document, we shall look primarily at the analysis of financial timeseries data, in particular the prices of stocks. For illustrative purposes, most of our examples will be drawn from the analysis of the price history of just a few stocks, as shown in Figure 2.1. Of course, in practice one must collect and analyse a large amount of data for a wide variety of stocks, in order to obtain more general and robust predictive models.

Chapter 2

Basic Time-Series Modelling

2.1 Predictive Model

Consider the stock prices shown in Figure 2.1, which vary daily over a period of time. Each sequence of prices for a given stock forms a time-series. Here let $V(t)$ be the price of a given stock at time t. In general, $V(t)$ represents the time-varying value of any asset. We consider only the gross, not net, asset value, and hence may assume that $V(t) \geq 0$.

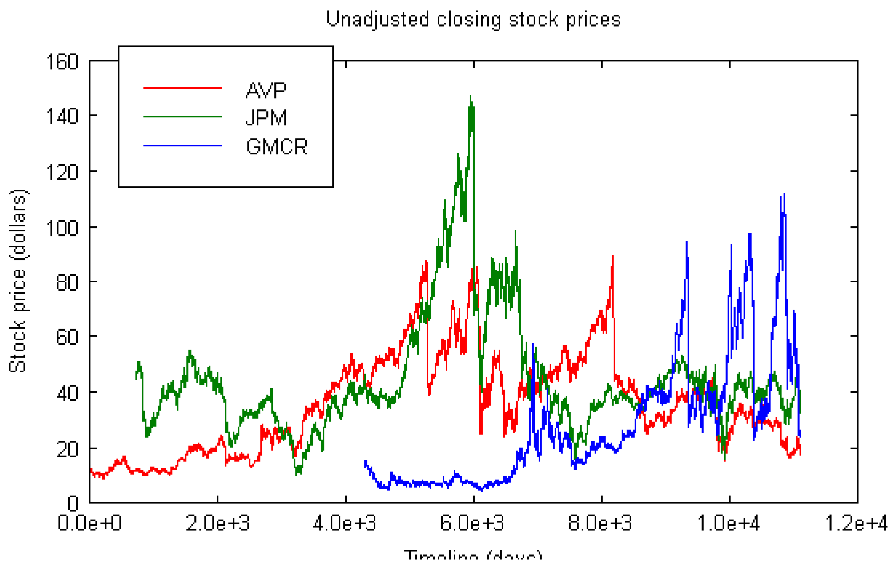

Figure 2.1: Example data showing the unadjusted closing prices of three different types of stock.

Now, suppose for the purpose of illustration that we buy an asset, e.g. a parcel of shares in a stock, at time t for amount $V(t)$ and then sell it at a later time $t+T$ for value $V(t+T)$. Then the difference in asset values is the *profit*, or *return*, on our investment, namely

$$P(t, t+T) = V(t+T) - V(t)\,, \tag{2.1}$$

as shown in Figure 2.2. Of key importance, therefore, is to attempt to define a predictive model that estimates the future value $V(t+T)$ given the current value $V(t)$.

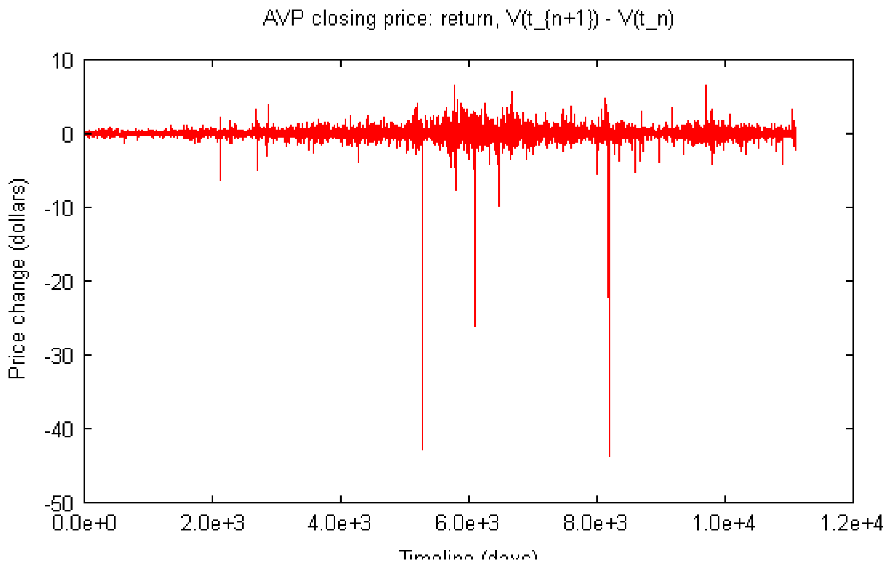

Figure 2.2: Example time-series of the daily return in closing price of AVP stock.

Now, observe that the return has the same currency units as the asset value. Thus, as an alternative, we might instead consider the profit per unit of investment, which is a dimensionless measure known as the *relative return*, given by

$$Q(t, t+T) = \frac{V(t+T) - V(t)}{V(t)}\,. \tag{2.2}$$

An example of the relative return is shown in Figure 2.3. Of more use, however, is the relative return on the investment per unit time, given by

$$R(t, t+T) = \frac{V(t+T) - V(t)}{T\,V(t)}\,. \tag{2.3}$$

This is known as the *rate of return*, and is show in Figure 2.4.

Next, observe that as the length T of the predictive interval $[t, t+T]$ becomes vanishingly small, we obtain the *instantaneous rate of return* given by

$$r(t) = \lim_{T\to 0} \frac{V(t+T) - V(t)}{T\,V(t)} = \frac{V'(t)}{V(t)}\,. \tag{2.4}$$

In order to relate $r(t)$ to a predictive model of $V(t+T)$, note from the derivative chain rule that

$$\frac{d\ \ln V(t)}{dt} = \frac{V'(t)}{V(t)} = r(t)\,. \tag{2.5}$$

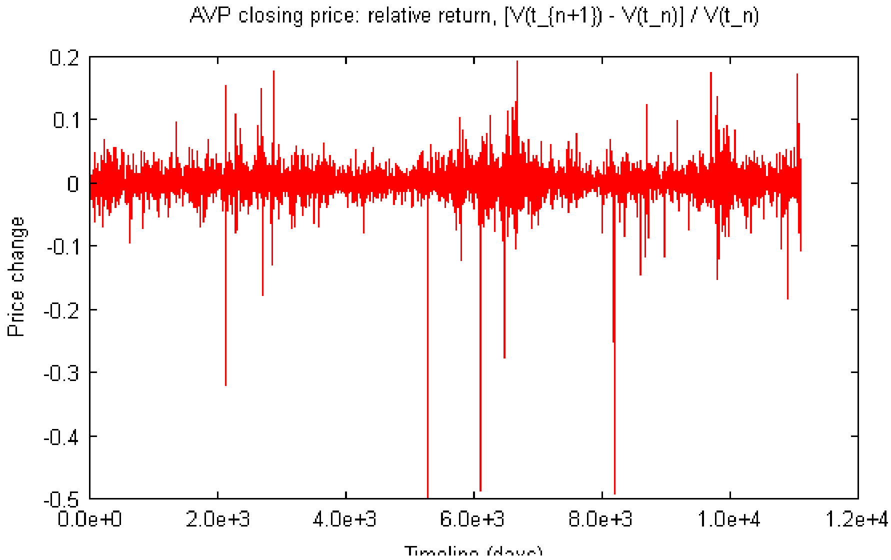

Figure 2.3: Example time-series of the daily relative return of the closing price of AVP stock.

Hence, by integrating this instantaneous rate of return over the interval $[t, t+T]$ we obtain

$$\ln V(t+T) \quad = \quad \ln V(t) + T\,\bar{R}(t, t+T)\,, \tag{2.6}$$

where $\bar{R}(t, t+T)$ is the *average* rate of return over the interval, given by

$$\bar{R}(t, t+T) \quad = \quad \frac{1}{T}\int_t^{t+T} r(t)\;dt\,. \tag{2.7}$$

Note that $\bar{R}(t, t+T)$ has the useful property of having the unrestricted range $(-\infty, \infty)$, despite the fact that $V(t)$ has the restricted range $[0, \infty)$.

Finally, observe for the sake of completeness that equation (2.6) can be rearranged to give

$$\bar{R}(t, t+T) \quad = \quad \frac{1}{T}\ln\frac{V(t+T)}{V(t)}\,, \tag{2.8}$$

which is thus also known as the *logarithmic* rate of return. Likewise, equation (2.6) can be rearranged to give

$$V(t+T) \quad = \quad V(t)\,e^{T\,\bar{R}(t,t+T)}\,, \tag{2.9}$$

such that $\bar{R}(t, t+T)$ is also known as the *continuously compounded* rate of return, due to the fact that

$$e^{T\,\bar{R}(t,t+T)} \quad = \quad \lim_{N\rightarrow\infty}\left(1+\frac{T\,\bar{R}(t,t+T)}{N}\right)^N\,. \tag{2.10}$$

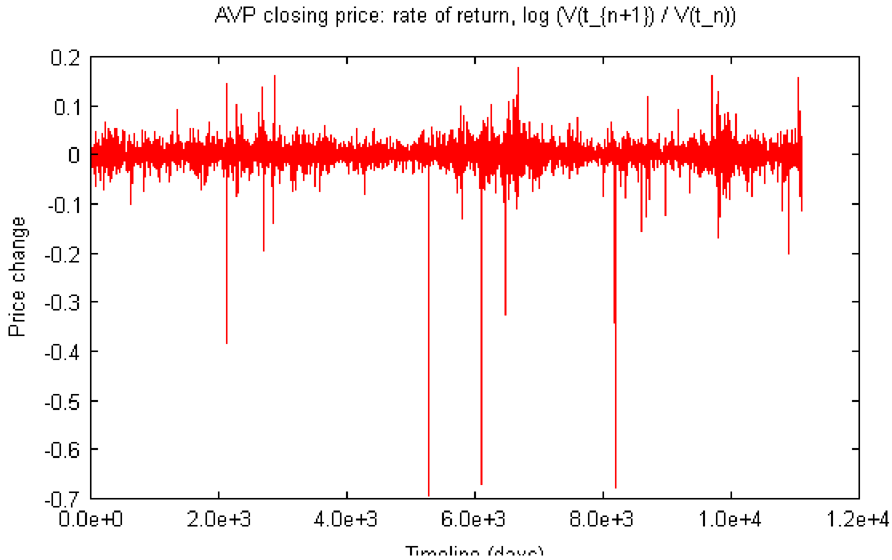

Figure 2.4: Example time-series of the daily rate of return of the closing price of AVP stock.

2.2 Stochastic Model

In practice, the average rate of return is often treated synonymously with the instantaneous rate of return. This is because, from the Mean Value Theorem (for example, see equation (3.1) from Appendix 3.1), there is a value $xi \in [t, t+T]$ such that $\bar{R}(t, t+T) = r(\xi)$. In other words, the averate rate of return over an interval equals the instantaneous rate of return at some point in the interval.

Since the value of ξ is unknown, we may treat it as a random variable, and by extension $\bar{R}(t, t+T)$ is also a random variable. It remains to decide how to model its stochastic distribution. In order to do this, first reconsider the stock prices shown in Figure 2.1. Although the value $V(t)$ of an asset is assumed to vary continuously with time t, the example stock prices were in fact sampled at the close of each day of trading. Thus, in general, rather than knowing $V(t)$ for all t, we only know values $V(t_i)$ sampled at discrete times t_i, for $i = 0, 1, \ldots, N$. The length of time $\delta t_i = t_i - t_{i-1}$ between samples is sometimes known as a *bar*. For the example stock prices, each bar is fixed equal to 1 day.

In terms of the prediction interval $[t, t+T]$, let us now arbitrarily define $t_0 = t$ and $t_n = t+T$ for some integer $n > 0$. Then $[t, t+T]$ may be partitioned into n subintervals $[t_{i-1}, t_i]$ for $i = 1, 2, \ldots, n$. Then, from definition (2.7), we obtain

$$\begin{aligned}(t_n - t_0)\bar{R}(t_0, t_n) &= (t_1 - t_0)\bar{R}(t_0, t_1) + \cdots + (t_n - t_{n-1})\bar{R}(t_{n-1}, t_n) \\ &= \sum_{i=1}^{n} \delta t_i \; \bar{R}(t_{i-1}, t_i)\,, \qquad (2.11)\end{aligned}$$

where $\bar{R}(t_{i-1}, t_i)$ is just the average rate of return over the subinterval $[t_{i-1}, t_i]$. Hence, we treat each $\bar{R}(t_{i-1}, t_i)$ as a random variable in its own right.

Turning briefly to sampling theory, note that if a random variable X is distributed with mean μ and variance σ^2, then the mean $\bar{X}$ of n random samples, each indpendently and identically drawn from the same distribution for X, is itself distributed with mean μ and variance σ^2/n. We therefore invoke several key assumptions:

1. *Independence*: Each variable $\bar{R}(t_{i-1}, t_i)$ is independently distributed from any other variable $\bar{R}(t_{j-1}, t_j)$ for $j \neq i$.
2. *Memorylessness*: The distribution of $\bar{R}(t_i, t_{i-1})$ depends only upon the length $\delta t_i = t_i - t_{i-1}$ of the subinterval $[t_{i-1}, t_i]$ and not upon its endpoints.
3. *Identity*: Any two distinct variables $\bar{R}(t_i, t_{i-1})$ and $\bar{R}(t_{j-1}, t_j)$, for $j \neq i$, are identically distributed if $\delta t_i = \delta t_j$.

As a consequence of the assumptions of memorylessness and identity, we may suppose that

$$\mathrm{E}[\bar{R}(t_i, t_{i-1})] \;=\; m(\delta t_i)\,\mu, \qquad \mathrm{Var}[\bar{R}(t_i, t_{i-1})] \;=\; v(\delta t_i)\,\sigma^2\,, \tag{2.12}$$

for unknown functions $m(\delta t)$ and $v(\delta t)$. Thus, taking the expectation of equation (2.11) gives

$$(t_n - t_0)\; m(t_n - t_0)\;\mu \;=\; \sum_{i=1}^{n} \delta t_i\; m(\delta t_i)\;\mu\,, \tag{2.13}$$

which is satisfied by choosing $m(\delta t) = 1$. Similarly, making use of the independence assmuption, taking the variance of eqaution (2.11) gives

$$(t_n - t_0)^2\; v(t_n - t_0)\;\sigma^2 \;=\; \sum_{i=1}^{n} (\delta t_i)^2\; v(\delta t_i)\;\sigma^2\,, \tag{2.14}$$

which is satisfied by choosing $v(\delta t) = 1/\delta t$. Therefore, we conclude that, over the arbitrary interval $[t, t+T]$, the continuous analogue to the discrete sample mean is given by

$$\mathrm{E}[\bar{R}(t, t+T)] \;=\; \mu, \qquad \mathrm{Var}[\bar{R}(t, t+T)] \;=\; \frac{\sigma^2}{T}\,. \tag{2.15}$$

2.3 Relative Returns

Suppose, for the sake of argument, that we invest an initial amount $V(t_0)$ in a bank that offers an annual rate of interest compunded n times a year. Then, letting S_i be the relative return over the interval $t \in [t_0 + (i-1)\delta t, t_0 + i\delta t]$, the value of our investment after one year is given by

$$V(t_0 + n\delta t) \;=\; V(t_0) \prod_{i=1}^{n} (1 + S_i)\,. \tag{2.16}$$

Next, suppose that the interest rate is fixed, such that each S_i is constant. Then equation (2.16) becomes

$$V(t_0 + n\delta t) \;=\; V(t_0)(1 + \bar{S})^n \;\text{ for } S_i \text{ constant}\,, \tag{2.17}$$

where $\bar{S}$ is the arithmentic mean of n values of some generic relative return S, defined by

$$1+\bar{S} \quad = \quad A_n(1+S) \;=\; \frac{1}{n}\sum_{i=1}^{n}(1+S_i)\,. \tag{2.18}$$

In contrast, suppose instead that the interest rate is now variable, such that S_i is no longer constant. Then equation (2.16) becomes

$$V(t_0+n\delta t) \quad = \quad V(t_0)(1+\tilde{S})^n \;\text{ for } S_i \text{ variable}\,, \tag{2.19}$$

where $\tilde{S}$ is the geometrically-adjusted mean relative return, defined by

$$1+\tilde{S} \quad = \quad G_n(1+S) \;=\; \left[\prod_{i=1}^{n}(1+S_i)\right]^{\frac{1}{n}}\,. \tag{2.20}$$

Now, it is known that for any variable X, we have $G_n(X) \leq A_n(X)$, where the equality occurs only when the values of X are constant. Hence, from equations (2.17) and (2.19), we observe that

$$V(t_0+n\delta t) \quad = \quad V(t_0)(1+\tilde{S})^n \;\leq\; V(t_0)(1+\bar{S})^n\,, \tag{2.21}$$

with the conclusion that the arithmetic mean relative return $\bar{S}$ over-estimates the geometrially-adjusted relative return $\tilde{S}$.

Suppose now that we do not know the precise value of each S_i, but we do know that each S_i is identically and independently sampled via the random variable S, which is a distributed with mean μ and variance σ^2. Next, observe from equation (2.22 that

$$\ln G_n(1+S) \quad = \quad \frac{1}{n}\sum_{i=1}^{n}\ln(1+S_i) \;=\; A_n(\ln(1+S))\,. \tag{2.22}$$

Chapter 3

Appendix

3.1 Mean Value Theorem

The Mean Value Theorem states that the average value of a function $f(x)$ over the interval $x \in [a, b]$ is given by $f(\xi)$ at some point $\xi \in [a, b]$. Mathematically, if $f(x)$ is an integrable function then the theorem implies that

$$f(\xi) \;=\; \frac{1}{b-a}\int_a^b f(x)\,dx \quad\Leftrightarrow\quad \int_a^b f(x)\,dx \;=\; (b-a)f(\xi)\,. \tag{3.1}$$

Graphically, the latter relation implies that the area under the curve over an interval is equal to a rectangular area on the same interval, where the height of the rectangle is just the mean value $f(\xi)$. Alternatively, since $f(x)$ is integrable, consider its indefinite integral

$$\int f(x)\,dx \;=\; F(x) + c\,,$$

where c is an arbitrary constant. Then the theorem states that

$$F(b) - F(a) \;=\; (b-a)\,F'(\xi)\,, \tag{3.2}$$

where $F'(x) = f(x)$.