ESTIMATING INFORMATION FROM STOCK PRICES IN THE CAPITAL MARKET VIA A NOVEL TREND

Asoka Korale, Ph.D.
School of Postgraduate Studies and Research
Sri Lanka Technological Campus
Padukka
asokak@sltc.ac.lk

***Abstract*—The paper presents a method to segment a time series by developing an estimate for the trend between two points in time and an estimate of the variation in the difference between the two points as a variation around this trend. The trend is estimated between two points in the time series and relates the amount of linear change between the points to the number of observations between them.**

In this paper, a time series of stock prices is used to estimate and detect changes in the trend. The variation in the prices provide information about the uncertainty in the decisions taken by the investors and the risks they faced at the time.

The model provides the means to estimate future values in the time series and control the uncertainty or risk in that estimate measured as the variation around the trend by choosing the interval or time period over which to make the estimate.

This model of a time series and trend is used to estimate the amount of information conveyed by the prices measured in an interval as the variation in the prices in that segment.

The period over which a future value in the time series is estimated determines the accuracy in its estimate and the variation or uncertainty in the estimate leading to the observation that future values are best estimated over periods where past trends can be extended or over periods where future behavior is expected to be similar to the past.

Keywords-Information; risk; random walk; insert (key words)

I. INTRODUCTION

Trends and patterns are observed in stock prices depending on how such trends and patterns are represented and detected. Trends in the prices convey important information about the security and are evidence of trends in the expectations of the investors that trade in them.

There is risk associated with such investments as the outcomes are not deterministic and are uncertain. The degree of uncertainty grows with the time interval over which the investment is realized as longer time periods allow greater variation in those outcomes as factors that influence those outcomes are more likely to change. The changes are also likely to be more significant as they are more uncertain or unpredictable the further into the future the return on the investment is realized.

Trends in the prices of these quantities influence the investors' expectations about the performance of the company and the return on an investment realized in the future.

Estimating trends in prices is important to reduce the uncertainty or risk associated with an investment realized in the future.

Trends are observed in quantities that vary. They may be quantities that vary over time and have a time dimension. A set of observations ordered in time can be considered a time series. Periods of time are also represented and defined by the trends that are observed during an interval in time.

II. ESTMATING TRENDS IN TIME SERIES

Trends are observed in time series as rising troughs and peaks or falling peaks and troughs [1]. We develop a method to detect the trends in the time series by accumulating the changes produced by runs of increments and decrements.

To illustrate the method, a random walk P_n is simulated as a sum of random variables. The random variables are independent and identically distributed (IID) and take values +1 and -1 with probability 0.5.

Equation (1) gives the incremental step of the random walk.

$$P_{n+1} = P_n + X_{n+1} \quad (1)$$

$$P_{n+2} = P_{n+1} + X_{n+2} = P_n + X_{n+1} + X_{n+2} \quad (2)$$

The series N steps apart is the sum of N IID steps.

$$P_{n+N} - P_n = \sum_{i=1}^{N} X_{n+i} \quad (3)$$

The distribution of this random variable can be estimated from the distribution of the first differences of the time series. Equation (4) defines the first differences on two data points in the time series. The random variable X_n is estimated over an interval N data points apart.

$$P_n - P_{n-1} = X_n \quad (4)$$

$$X_n \sim F(\mu_x, \sigma_x) \quad (5)$$

The expected value of the random variable X_n can be estimated as the expected value of the first differences of the series P_n.

$$E(P_n - P_{n-1}) = E(X_n) = \mu_x \quad (6)$$

The variance of the random variable X_n

$$Var(P_n - P_{n-1}) = Var(X_n) = \sigma_x^2 \quad (7)$$

The estimate for the expected value of a change in the series N_s steps apart is

$$E(P_{n+N_s} - P_n) = \mathrm{E}\{\sum_{i=1}^{N_s} X_{n+i}\} = N_s \mu_x \quad (8)$$

The variance of this difference in prices is a sum of variances and pairwise covariances of the random variables X_{n+i}, i = $\{1,2 \ldots, N_s\}$.

$$Var(P_{n+N_s} - P_n) = \sum_{i=1}^{N_s} Var(X_{n+i}) + \sum_{\forall i,j,\ i\neq j} Cov(X_i, X_j) \quad (9)$$

The covariances are zero when the random variables are IID. The variance in the series N_s steps apart is given by equation (10)

$$Var(P_{n+N_s} - P_n) = \sum_{i=1}^{N_s} Var(X_{n+i}) = N_s \sigma_x^2 \quad (10)$$

Future values in this series can be estimated by extending the trend established by the difference N observations apart.

$$E(P_{n+N+1} - P_n) = (N+1)\mu_x \quad (11)$$

$$P_{n+N+1} = P_n + (N+1)\mu_x \quad (12)$$

This estimate can be made when the random variable is estimated over the interval defined by the difference an interval N observations apart.

A. A random walk example of a random process

A random walk example is used to illustrate the methods developed to identify trends, changes in trends, risks, entropy and to estimate future values in the series.

Figure1, depicts a random walk. The random walk exhibits trends. The series diverges with the number of steps, n. The trends too will exhibit this behavior.

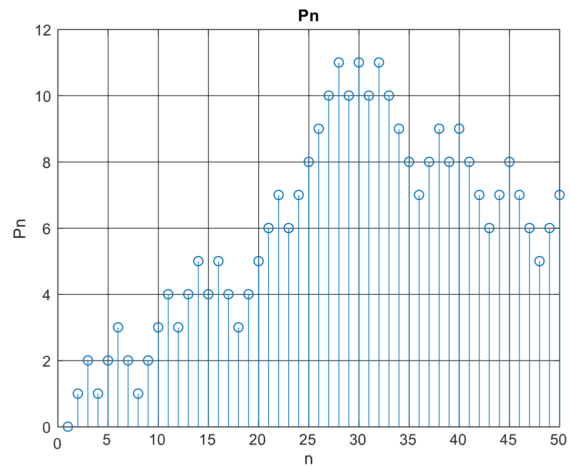

Figure 1. Random walk example of a random process

The random walk used to generate the time series has a step size X_n of value 1 or -1 with probability 0.5. The expected value of a step $E(X_n)$ given by equation (13)

$$E(X_n) = (X_n = 1)(1) + \mathrm{P}(X_n = -1)(-1) = 0 \quad (13)$$

The variance $Var(X_n)$ associated with this step is 1 given by

$$Var(X_n) = \mathrm{P}(X_n = 1)(1 - E(X_n))^2 + \mathrm{P}(X_n = 1)(-1 - E(X_n))^2 = 1 \quad (14)$$

B. Interval of estimaton

The interval over which to estimate future values in the time series depends on the random process that is used to model the observations.

In the case of a random walk, the time series can be modeled as a sum of random variables.

The series N steps apart is the sum of N differences. This difference in the series can be represented by equation (15).

$$P_{n+N} - P_n = \sum_{i=1}^{N} X_{n+i} \quad (15)$$

The mean of the first difference is given by equation

$$E\{X_{n+1}, X_{n+2}, \ldots, X_{n+\mathrm{N}}\} = \mu_{dx} \quad (16)$$

The degree to which each first difference in the series deviates from its mean can be estimated from equation (17). It is estimated over the interval of the trend.

It is a way of measuring how well a linear trend fits the data points. Lower values of V_N provide a better fit.

$$V_N = \frac{1}{N}\sum_{i=1}^{N} \left(\frac{X_{n+i}}{\mu_{dx}} - 1\right)^2 \quad (17)$$

The ratio of each change in price to the average change in price measured over the interval of the trend is a measure of the average return at each price change.

$$r = \frac{X_{n+i}}{\mu_{dx}} \quad (18)$$

III. INFORMATION IN THE TIME SERIES

Stocks are usually traded on prices and the variation in the prices convey the risks associated with an investment that may rise or fall in value as the variation is uncertain.

The variation in the prices can be measured by its variance. It is a measure of the risk associated with it.

There are trends and seasonality in the prices depending on economic, political and social conditions. Estimates of these trends and cycles provide information about the economy and society.

A. Information provided by trends

Trends are found in time series and indicate intervals over which the data points rise or fall linearly. The degree to which this rise or fall is linear can be estimated by considering the first differences X_n in the time series, the random process underlying modeled as a random walk.

The variation in prices can be measured over a trend by measuring the variance over the period of the trend. It can be a measure of the information conveyed by the trend. It can also be used to compare between trends.

B. Models of trends [2]

Time series of stock prices are modeled as a time dependent trend and a variation around it that depends on the interval of time between successive values

In this model the uncertainty in the time dependent trend depends on a random variable drawn from a Normal distribution with variance that depends on the interval of time between two prices.

A stochastic process X can be modeled a Newtonian term based on dt and a Brownian term based in the infinitesimal increment of W which is called dW_t . The Brownian term of X can have a noise σ_t, and the infinitesimal change of X_t is given by equation (19).

$$dX_t = \sigma_t dW_t + \mu_t dt \quad (19)$$

The drift μ_t can be dependent on time t.

σ_t is the volatility of the process X at time t and μ_t the drift of X at t.

The Brownian motion process $W = (dW_t : t \geq 0)$ is P-Brownian motion if and only if

1. W_t is continuous, and $W_0 = 0$
2. The value of W_t is distributed, under P, as a normal random variable N(0,t)
3. The increment W_{s+t} - W_t is distributed as Normal N(0,t), under P, and is independent of F_s, the history of what the process did up to time s.

The parameters of this model are estimated from historical prices.

IV. RESULTS ON EXTENDING TRENDS

A time series of daily oil prices is used to demonstrate and validate the methods developed to detect trends and changes in trends. It is also used to estimate the variation in the prices over a period defined by a trend as a way to estimate the information in it. The uncertainty in the random variables estimated from the prices is used to estimate the uncertainty in the estimate of future values obtained by extending the trend.

A. Validation of detected trends and patterns

A time series of daily oil prices depicted in Figure (2) is used to validate the models developed. There are approximately 250 trading days in a year and the time series depicted has 2448 data points. The 10-year period from January 2010 to January 2020 show peaks, troughs and variations in prices depending on political, economic and social conditions that influence the demand and supply for oil.

The changes in prices observed can be attributed to these factors. Oil is traded on the information available to the investors and their expectations of future conditions as investments are realized over a period of time. This time period is measured in daily intervals and future values of this time series can be estimated by extending past trends.

The accuracy in estimating future values depends on the ability to estimate past trends and the likelihood that such trends will persist or extend into the future. The variations in the prices are attributed to the uncertainty in those conditions that investors have at the time of trading. The prices are also influenced by information available to the public ensuring an efficient market.

The oil prices plotted as continuous curve over the 10-year period in Figure 2 is plotted as a stem plot in Figure 3. An interval of 100 days is selected from the 10-year period from this plot and conform to the data points in the interval [201,300] depicted in Figure 3.

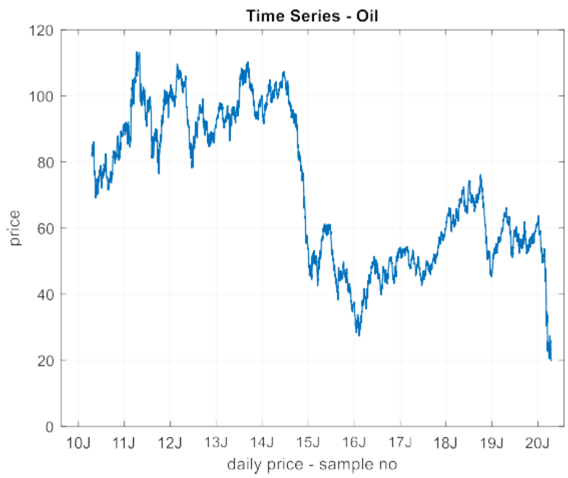

Figure 2. Oil prices (USD) in the interval [1,2048]

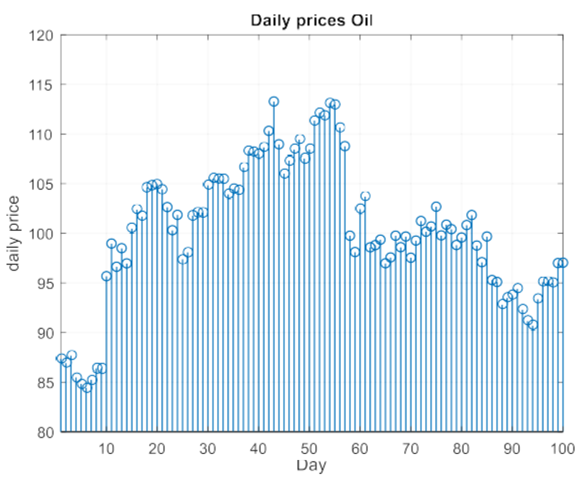

Figure 3. Oil prices (USD) in the itnerval [201,300]

B. Interval of estimaton

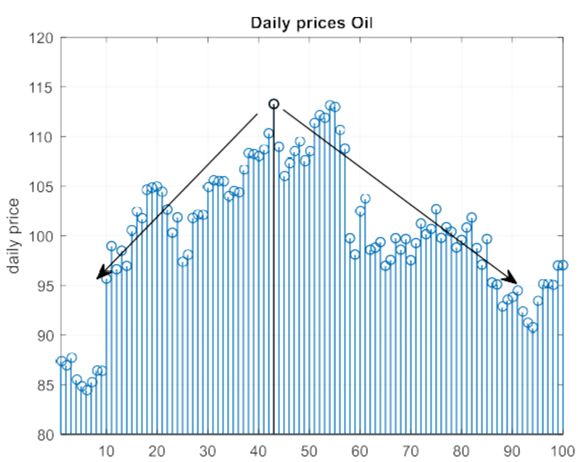

Figure 4. Oil prices (USD) in the itnerval [201,300]

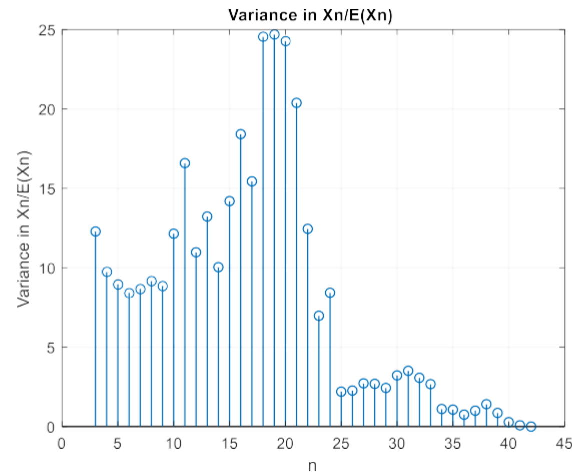

Figure 5. The measure of Vn over interval [43,3]

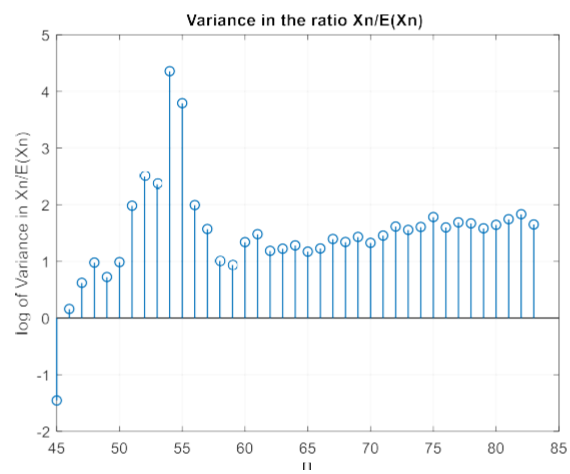

Figure 6. The measure of Vn over interval [43,84]

Figures 5 and 6 depict how well a linear trend established between two points fits the data points by evaluating V_N from the peak at n = 43 in backward and forward directions in time respectively. Figure 4 depicts the direction of evaluation Vn.

Figure 5 plots the measure V_N as a function of decreasing n evaluting the trend established on the intervals [n2, n1] where n2 = 43 and n1 = [42, 41, …, 3].

The interval over which to estimate the trend depends on the future value to be estimated. The future value is estimated by extending a trend estimated on past data. It is a number of data points outside the interval used to estimate the trend. The error in the estimate and the uncertainty in the prediction depend on the past values used to estimate the future value.

The interval to estimate $P_n(44)$ from past data depends on the behavior of the trend extended from the past beginning from $P_n(43)$. Trend segments could then be defined between

[43,42], [43,41], $[43,43-N]$ where N is a number of data points N forming a window of length N.

The sequence of V_N depicted in Figure 5, are changes that can be used to detect those intervals. The interval [43,25] and [43,9] are such intervals where the past data fits the trend well. The data point P_n (44) however may not lie on the trend extended over historical data as the changes in prices are expected to be random and follow a random walk.

C. Validate estimates of future values by extending a trend line

Future values in the time series are estimated by extending the trend line from historical data. Figure 7 depicts a trend extended between P_n (43) and the data points preceding it. This is a forecast one day ahead using trends extended from the previous day into the past.

The variation in the error in the estimate of P_n (44) from a trend extended between P_n (43) and each of the preceding data points is plotted in Figure 8.

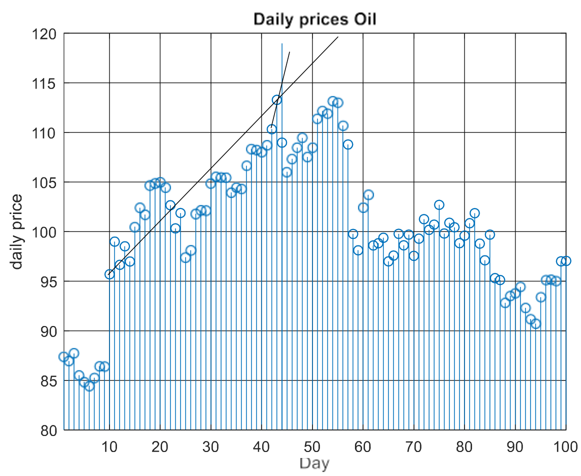

Figure 7. Trends extended on oil prices depicted in Figure 15

Figure 7 depicts how well the two trend lines fit the data. This is measured as ratio of each price change to the average change in the prices over the trend via the measure defined by V_N given by equation (17).

The variance in the random variable X_n provides an estimate in the variance of a change in the prices. In this case it is estimated over the interval of the trend. The variation in X_n measures how much each step X_i in the interval [n1, n2] deviates from its average or mean value $E\{X_i\}$ measured over the same interval.

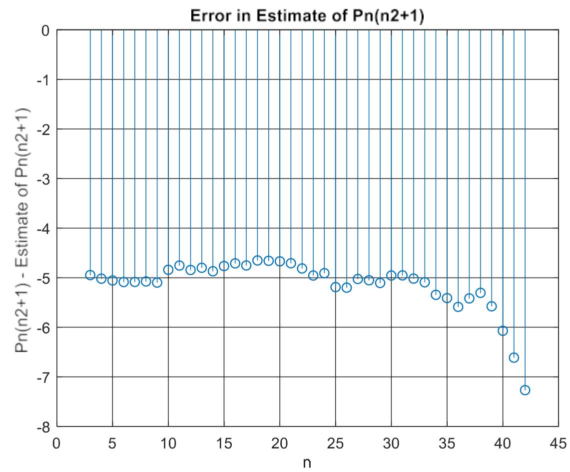

Figure 8. Error in estimating Pn(n2+1) by extending trend

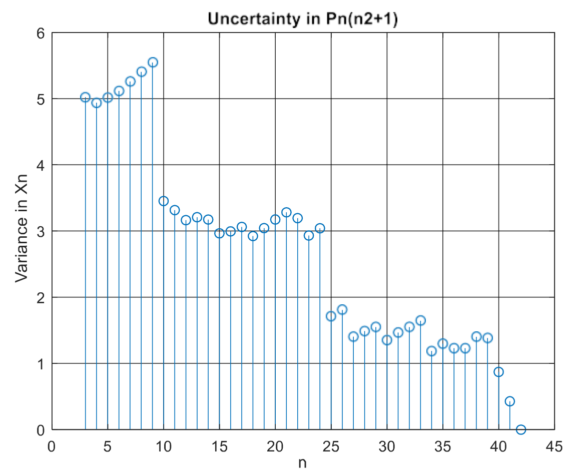

Figure 9. Uncertainty in the error in estimating Pn(44) [43,:]

TABLE I. VARIABLES USED IN TABLE 1

n	data point number
E{Xn}	Expected value of Xn estimated over the interval [n2,n1]
Var{Xn}	Variance of Xn estimated over the interval [n2,n1]
E{Pn}	Expected value of Pn estimated over the interval [n2,n1]
Var{Pn}	Variance of Pn estimated over the interval [n2,n1]
PN	Estimate of Pn(n2+1) by extending trend over [n2,n1]
Error PN	Error in the estimate of Pn(n2+1) , Pn(n2+1) - PN
Vn	Measure to estimate fit of random variable to trend

TABLE II. ANALYSIS OF TRENDS IN INTERVAL [43,4]

n	E{Xn}	Var{Xn}	E{Pn}	Var{Pn}	PN	PeN	Vn
42	2.96	0.00	111.80	2.19	116.24	-7.27	0.00
41	2.31	0.43	110.76	3.64	115.59	-6.62	0.08
40	1.76	0.87	110.07	4.16	115.04	-6.07	0.28
39	1.27	1.38	109.69	3.89	114.55	-5.58	0.86
38	1.00	1.40	109.46	3.51	114.28	-5.31	1.41
37	1.11	1.23	109.05	4.00	114.39	-5.42	1.00
36	1.28	1.23	108.46	5.96	114.56	-5.59	0.75
35	1.10	1.30	108.01	6.88	114.38	-5.41	1.07
34	1.04	1.19	107.61	7.69	114.32	-5.35	1.11
33	0.79	1.65	107.41	7.38	114.07	-5.10	2.67
32	0.71	1.55	107.25	7.06	113.99	-5.02	3.07
31	0.65	1.47	107.11	6.73	113.93	-4.96	3.51
30	0.65	1.35	106.95	6.59	113.93	-4.96	3.22
29	0.80	1.55	106.63	7.61	114.08	-5.11	2.43
28	0.74	1.49	106.35	8.33	114.02	-5.05	2.69
27	0.72	1.41	106.08	8.99	114.00	-5.03	2.72
26	0.89	1.81	105.64	11.83	114.17	-5.20	2.27
25	0.88	1.71	105.20	14.59	114.16	-5.19	2.20
24	0.60	3.04	105.03	14.39	113.88	-4.91	8.42
23	0.65	2.93	104.81	14.72	113.93	-4.96	6.98
22	0.51	3.20	104.71	14.25	113.79	-4.82	12.45
21	0.40	3.28	104.70	13.64	113.68	-4.71	20.38
20	0.36	3.18	104.71	13.07	113.64	-4.67	24.27
19	0.35	3.05	104.72	12.55	113.63	-4.66	24.69
18	0.35	2.92	104.71	12.07	113.63	-4.66	24.54
17	0.45	3.06	104.60	11.94	113.73	-4.76	15.44
16	0.40	3.00	104.52	11.69	113.68	-4.71	18.42
15	0.46	2.97	104.38	11.83	113.74	-4.77	14.20
14	0.56	3.17	104.14	13.20	113.84	-4.87	10.04
13	0.49	3.21	103.95	13.77	113.77	-4.80	13.23
12	0.54	3.17	103.73	14.96	113.82	-4.85	10.97
11	0.45	3.32	103.58	15.17	113.73	-4.76	16.59
10	0.53	3.45	103.35	16.51	113.81	-4.84	12.14
9	0.79	5.55	102.86	24.03	114.07	-5.10	8.86
8	0.77	5.41	102.41	30.68	114.05	-5.08	9.17
7	0.78	5.26	101.94	37.60	114.06	-5.09	8.68
6	0.78	5.12	101.48	44.46	114.06	-5.09	8.42
5	0.75	5.02	101.06	50.24	114.03	-5.06	8.96
4	0.71	4.94	100.67	54.88	113.99	-5.02	9.74

Table II, can be used to calculate statistics on the variation the estimates and the random variables including estimating the distributions on it.

Table II, shows that the variance in the random variable tends to increase with the length of the interval used to estimate the trend. It is a measure of the variation in the changes in the prices observed over that interval.

The total variation in the price difference between prices P_{n+N} and $P_n$is the variation of a sum of random variables

$$Var(P_{n+N} - P_n) = \sum_{i=1}^{N} Var(X_{n+i}) + \sum_{\substack{\forall i,j \\ i\neq j}} Cov(X_i, X_j)$$

(20)

V. CONCLUSIONS

The paper developed methods to estimate linear trends in time series. The trend is defined as a difference in two prices an interval apart.

The trends are detected at different scale depending on how well the trend fits the data. The trend is estimated over a number of data points sampled uniformly.

The linear trend can be assessed in relation to how well it fits the data points by measuring how close the ratio of each change in price to its mean deviates from unity. The interval over which to estimate the trend can be estimated by observing the change in this quantity over time.

The linear trend can be extended to estimate future values in the time series. The uncertainty in this estimate can also be estimated by the variation in the price changes measured over the interval of the trend.

Since the trend is estimated over different time scales or a number of samples used to define an interval for the trend, forecasts can be made over different time intervals where the future trends are expected to resemble the past.

The trends in the prices convey information about events that influence them. The changes in trend of different scale can be used to detect changes in those factors that produce the variation in the prices.

The variation of the prices over the interval of the trend can be used to estimate the risks that investors faced at the time of trading.

REFERENCES

[1] M. Baxter, and A. Rennie, “Financial Calculus”, An introduction to derivative pricing, Cambridge Univerisity Press, pp. 48–57, 1996.

[2] John J. Murphy, Technical analyis of financial markets, a comprehensive guide to trading methods and applications, 1999