

## The general structure of market data

*By Dr David Worrall*

The statistical analysis of trading data, and stochastic modelling using it, is an ongoing area of research in finance because, despite intensive study, a comprehensive understanding of the structure of capital markets exchange trading data remains elusive. For an overview of this issue, Chapter 5 of [Sonification and Information](#) provides a more formal introduction, including more detail of the techniques illustrated here.

As market-time unfolds, the price of a traded security ‘wanders’ from value to value. This is known mathematically as a stochastic process, or a random function. Amongst the simplest types of such stochastic processes is a random walk, called so because it may be thought of as a simplified model for an individual walking on a line who at each point of time either takes one step to the right with probability  $p$  or one step to the left with probability  $(1-p)$ . Because the domain over which this function is defined is a time interval, and prices move discretely (step from one value to another), this random function is also known as a discrete time series and securities trading datasets can thus be classified as discrete multidimensional time series whose moment-to-moment behaviour is analogous to geometric Brownian motion.

Two principal concerns are ways to accurately describe the way prices are distributed statistically and whether or not, and to what extent, auto-correlation exists and can be detected, even pre-empted, as market prices evolve. Understanding the accuracy of the description of price distribution is important for the risk analysis of various trading instruments in the longer term, and understanding the inherent autocorrelation is important in attempts to predict future, especially catastrophic, events.

### **A little history: The Efficient Market Hypothesis**

In 1900, Louis Bachelier, using methods that had been created for analysing gambling, conjectured that price fluctuations in day-to-day exchange-traded securities were independent random variables. He provided little empirical evidence to support the assumption, yet his work, which also included techniques for analysing the value of government bonds and displaying options-related strategies, is today considered seminal. Bachelier’s thesis was revolutionary, but largely ignored and forgotten and it was Albert Einstein’s independent description that brought the solution for Brownian motion to the attention of physicists.

The unexpected crashes of stock markets in 1929, and the depression that followed, stimulated mathematicians to attempt a better understanding of market action through statistical analysis. In 1962 Paul Cootner published an anthology of quantitative analysis that became the basis for what is known as the Efficient Market Hypothesis, which Eugene Fama later formalised in 1965. Simply stated, the Efficient Market Hypothesis assumes that the market price of a trading security reflects all that is known about it; the difference in price from one point in time to another, simply reflecting any new information about the security as it becomes known. After digesting the new information together with an assessment of the risks involved, the collective consciousness that is the market finds an equilibrium price. This is called the random walk version of Efficient Market Hypothesis. A random walk is a sufficient but not a necessary condition for market efficiency. That is, whilst market efficiency does not necessarily imply a random walk (it may be some other process), a random walk does imply market efficiency. Such a random walk is called Brownian motion. If Bachelier’s conjecture is correct, as is assumed by the Efficient Market Hypothesis, prices would exhibit no autocorrelation and a statistical analysis would reveal a normally-distributed, or Gaussian, probability density function (PDF).

### **Mandelbrot ups the ante**

As a result of the discrepancy between this theoretical perspective and his analysis of the way certain speculative prices, such as those of cotton, moved, Benoit Mandelbrot became dissatisfied with this simplified model. Mandelbrot is one of the significant contributors in the field and his technical monograph summarises a unique and influential perspective built over many decades. His investigations showed that real markets exhibit much larger variability, as well as greater leptokurtosis and skewness than a normal distribution. He realised that the market process could be better described by a Lévy flight.

In addition to their distribution, markets also exhibit momentary autocorrelations. Modern econometrics and financial engineering place considerable import on understanding such phenomena because the increased likelihood of extreme events, both positive and negative, indicates greater market volatility than if the markets are normally distributed, which, in turn, would impact on risk assessment, options pricing and portfolio theory in general.

The principal statistical techniques used in market analysis and simulation today, use Bayesian partitioning, variable length Markov chains and Monte Carlo simulations, and Generalized Auto-Regressive Conditional Heteroskedasticity (GARCH) modelling. There seems nothing intrinsic in these techniques to limit the application of sonification to them and preliminary discussions indicated that some researchers in the field are interested in experimenting with such techniques to assist in the comprehension of their abstraction.