Dynamics and time series: theory and applications

Stefano Marmi
Scuola Normale Superiore
Lecture 9, Mar 19, 2009

- Lecture 1: An introduction to dynamical systems and to time series. Periodic and quasiperiodic motions. (Tue Jan 13, 2 pm - 4 pm Aula Bianchi)
- Lecture 2: Ergodicity. Uniform distribution of orbits. Return times. Kac inequality Mixing (Thu Jan 15, 2 pm - 4 pm Aula Dini)
- Lecture 3: Kolmogorov-Sinai entropy. Randomness and deterministic chaos. (Tue Jan 27, 2 pm - 4 pm Aula Bianchi)
- Lecture 4: Time series analysis and embedology. (Thu Jan 29, 2 pm - 4 pm Dini)
- Lecture 5: Fractals and multifractals. (Thu Feb 12, 2 pm - 4 pm Dini)
- Lecture 6: The rhythms of life. (Tue Feb 17, 2 pm - 4 pm Bianchi)
- Lecture 7: Financial time series. (Thu Feb 19, 2 pm - 4 pm Dini)
- Lecture 8: The efficient markets hypothesis. (Tue Mar 3, 2 pm - 4 pm Bianchi)
- Lecture 9: A random walk down Wall Street. (Thu Mar 19, 2 pm - 4 pm Dini)
- Lecture 10: TBA. (Tue Mar 24, **11 am– 1 pm** aula ?? Rosario Mantegna)

- Seminar I: Waiting times, recurrence times ergodicity and quasiperiodic dynamics (D.H. Kim, Suwon, Korea; Thu Jan 22, 2 pm - 4 pm Aula Dini)
- Seminar II: Symbolization of dynamics. Recurrence rates and entropy (S. Galatolo, Università di Pisa; Tue Feb 10, 2 pm - 4 pm Aula Bianchi)
- Seminar III: Heart Rate Variability: a statistical physics point of view (A. Facchini, Università di Siena; Tue Feb 24, 2 pm - 4 pm Aula Bianchi)
- Seminar IV: Study of a population model: the Yoccoz-Birkeland model (D. Papini, Università di Siena; Thu Feb 26, 2 pm - 4 pm Aula Dini)
- Seminar V: Scaling laws in economics (G. Bottazzi, Scuola Superiore Sant'Anna Pisa; Tue Mar 17, 2 pm - 4 pm Aula Bianchi)
- Seminar VI: Complexity, sequence distance and heart rate variability (M. Degli Esposti, Università di Bologna; Thu Mar 26, 2 pm - 4 pm Aula Dini)
- Seminar VII: Forecasting (M. Lippi, Università di Roma; late april, TBA)

Today's bibliography:

S.J. Taylor "Asset price dynamics, volatility and prediction" Princeton University Press, Ch. 5

K. Cuthbertson "Quantitative Financial Economics" Chapter 4 Valuation and 5 The efficient markets hypothesis

Eugene Fama: "Foundations of Finance" Chapter 5: Efficient Capital Markets

Tomas Björk: "Arbitrage Theory in Continuous Time" Chapter 4 and Appendix B

What is an efficent capital market?

An efficient capital market is a market which is efficient in processing information: the prices of securities observed at any time are based on “correct” evaluation of all information available at that time. Prices “fully reflect” available information. The prices are always “fair”, they are good indicators of value

The concept of market efficiency had been anticipated at the beginning of the century: Bachelier (1900) writes "past, present and even discounted future events are reflected in market price, but often show no apparent relation to price changes". And also "if the market, in effect, does not predict its fluctuations, it does assess them as being more or less likely, and this likelihood can be evaluated mathematically".

Weak vs. strong efficiency

More formally: a capital market is said to be efficient if it fully and correctly reflects all relevant information in determining security prices. Formally, the market is said to be efficient with respect to some information set, Θ_t , if security prices would be unaffected by revealing that information to all participants. Moreover, efficiency with respect to an information set, Θ_t, implies that it is impossible to make economic profits by trading on the basis of Θ_t.

The weak form of the efficient market hypothesis claims that prices fully reflect the information implicit in the sequence of **past prices**. The semi-strong form of the hypothesis asserts that prices reflect all relevant information that is **publicly** available, while the strong form of market efficiency asserts information that is known to **any** participant is reflected in market prices.

Fundamental vs. technical analysis

Fundamental analysis maintains that markets may misprice a security in the short run but that the "correct" price will eventually be reached. Analyzing financial statements, management and competitive advantages, one can accurately estimate a “fair value” for the stock. Profits can be made by trading the mispriced security and then waiting for the market to recognize its "mistake" and reprice the security.

Technical analysis maintains that all information is reflected already in the stock price, so fundamental analysis is a waste of time. Trends 'are your friend' and sentiment changes predate and predict trend changes. Investors' emotional responses to price movements lead to recognizable price chart patterns. Technical analysis does not care what the 'value' of a stock is. Their price predictions are only extrapolations from historical price patterns.

Strong and semi-strong efficiency

In the semi-strong form of the EMH a trading strategy incorporating current publicly available fundamental information (such as financial statements) and historical price information will not systematically outperform a buy-and-hold strategy. Share prices adjust instantaneously to publicly available new information, and no excess returns can be earned by using that information. Fundamental analysis will not be profitable.

In strong-form efficiency share prices reflect all information, public and private, fundamental and historical, and no one can earn excess returns. Inside information will not be profitable.

Weak, semi-strong and strong EMH

- **Weak EMH**. "One can not use *past price and volume values* to craft investment strategies that can reliably out perform the over all market."
- **Semi-Strong EMH**. "One cannot use *publically available information* to to craft investment strategies that can reliably out perform the over all market."
- **Strong EMH**. "One cannot use *any information* --- *including material, non-public information* --- to to craft investment strategies that can reliably out perform the over all market."

http://www-stat.wharton.upenn.edu/~steele/Courses/434/434index.html

J. Michael Steele course on Financial Time Series

Steele's comment on the EMH:

The second part of the agenda is to open the conversation about the Efficient Market Hypothesis (EMH). Everyone seems to know what this is all about until someone says --- "great, lets write this out as mathematics."

At that point almost everyone starts to become uneasy. The fact is that that there are some hard --- and debatable --- decisions must be made. Our first step will be to round up the "usual suspect." We'll then see what comes out in the wash. Incidentally, I have collected some sources on the EMH that may be useful to you, and you will surely have loads of material from your other courses.

A probability space is simply a measure space $(\Omega, \mathcal{F}, P)$ where the measure P has the property that it has total mass equal to unity, i.e.

$$P(\Omega) = 1.$$

The underlying space Ω often is referred to as the **sample space**, and the elements of the sigma-algebra $\mathcal{F}$ are called **events**.

B.1 Random Variables and Processes

In this section we will discuss random variables and random processes.

Definition B.1 *A* **random variable** X *is a mapping*

$$X : \Omega \to R$$

such that X *is* $\mathcal{F}$*-measurable.*

Definition B.2 *The* **distribution measure** μ_X *for a random variable* X *is a measure on* $(R, \mathcal{B})$ *defined by*

$$\mu_X(B) = P\left(\{\omega \in \Omega;\ X(\omega) \in B\}\right), \quad B \in \mathcal{B},$$

i.e.

$$\mu_X(B) = P(X^{-1}(B)).$$

The (cumulative) **distribution function** *of* X *is denoted by* F_X *and defined by*

$$F_X(x) = P\left(\{\omega \in \Omega;\ X(\omega) \leq x\}\right).$$

Definition B.3 *For any* $X \in L^1(\Omega, \mathcal{F}, P)$ *its* **expected value**, *denoted by* $E[X]$, *is defined by*

$$E[X] = \int_\Omega X(\omega)\, dP(\omega).$$

$$E[X] = \int_0^\infty P(X \geq t)\, dt.$$

For $X \in L^2$ *the* **variance** *is defined by*

$$\mathrm{Var}[X] = E\left[(X - E[X])^2\right].$$

Definition B.6 *A random process on the probability space* $(\Omega, \mathcal{F}, P)$ *is a mapping*

$$X : R_+ \times \Omega \to R,$$

such that for each $t \in R_+$ *the mapping*

$$X(t, \cdot) : \Omega \to R,$$

is $\mathcal{F}$*-measurable.*

Björk : "Arbitrage Theory in Continuous Time"

Definition B.11 *The sigma-algebra* $\sigma\{X\}$ *is defined as the smallest sigma-algebra* $\mathcal{F}$ *such that* X *is* $\mathcal{F}$*-measurable.*

We will refer to $\sigma\{X\}$ as "the sigma-algebra generated by X". Technically speaking it is the intersection of all sigma-algebras $\mathcal{G}$ such that X is $\mathcal{G}$-measurable, but we can in fact give a more explicit representation.

Proposition B.12 *We have the representation*

$$\sigma\{X\} = \left\{X^{-1}(B);\ B \in \mathcal{B}(R)\right\}.$$

Taylor, Asset Price Dynamics, Volatility and Prediction, P.U.P. (2005)

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2 \right). \tag{3.1}$$

This density has two parameters: the mean μ and the variance σ^2 of the random variable. We use the notation $X \sim N(\mu, \sigma^2)$ when X has the above density.

A linear function of a normal variable is also normal. If $X \sim N(\mu, \sigma^2)$ and $Y = a + bX$, then $Y \sim N(a + b\mu, b^2\sigma^2)$. In particular, with $a = -\mu/\sigma$ and $b = 1/\sigma$,

$$X \sim N(\mu, \sigma^2) \quad \Rightarrow \quad Z = \frac{X-\mu}{\sigma} \sim N(0, 1).$$

We call Z the *standard normal* distribution. Its d.f. is simply

$$f(z) = \frac{1}{\sqrt{2\pi}} \exp(-\tfrac{1}{2}z^2), \tag{3.2}$$

and we may denote its c.d.f. by $\Phi(z)$, which has to be evaluated by numerical methods. The probabilities of outcomes for X within particular ranges can be calculated from

$$P(a \leqslant X \leqslant b) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right).$$

The density of the normal distribution is symmetric about its mean μ. Symmetry ensures that all the odd central moments are zero and therefore the skewness of the distribution is zero. The second and fourth central moments are respectively σ^2 and $3\sigma^4$, so that all normal distributions have a kurtosis equal to three.

Exponential functions of normal variables are often encountered in finance. The general result for their expectations is

$$E[e^{uX}] = \exp(u\mu + \tfrac{1}{2}u^2\sigma^2). \tag{3.3}$$

Lognormal distribution

A random variable Y has a lognormal distribution whenever $\log(Y)$ has a normal distribution. When $\log(Y) \sim N(\mu, \sigma^2)$, the density function of Y is

$$f(y) = \begin{cases} \dfrac{1}{y\sigma\sqrt{2\pi}} \exp\left(-\dfrac{1}{2}\left(\dfrac{\log(y) - \mu}{\sigma} \right)^2 \right), & y > 0, \\ 0, & y \leqslant 0. \end{cases} \tag{3.4}$$

From equation (3.3), $E[Y^n] = \exp(n\mu + \frac{1}{2}n^2\sigma^2)$ for all n. Consequently, the mean and the variance of Y are

$$E[Y] = \exp(\mu + \tfrac{1}{2}\sigma^2) \quad \text{and} \quad \operatorname{var}(Y) = \exp(2\mu + \sigma^2)(\exp(\sigma^2) - 1).$$

The mean exceeds the median, namely $\exp(\mu)$, reflecting the positive skewness of this nonsymmetric distribution.

Taylor, Asset Price Dynamics, Volatility and Prediction, P.U.P. (2005)

Table 3.1. Definitions of ten types of stochastic process.

A process is...	*If...*
Strictly stationary	The multivariate distribution function for k consecutive variables does not depend on the time subscript attached to the first variable (any k).
Stationary	Means and variances do not depend on time subscripts, covariances depend only on the difference between the two subscripts.
Uncorrelated	The correlation between variables having different time subscripts is always 0.
Autocorrelated	It is not uncorrelated.
White noise	The variables are uncorrelated, stationary and have mean equal to 0.
Strict white noise	The variables are independent and have identical distributions whose mean is equal to 0.
A martingale	The expected value of variable t, conditional on the information provided by all previous values, equals variable $t-1$.
A martingale difference	The expected value of variable t, conditional on the information provided by all previous values, always equals 0.
Gaussian	All multivariate distributions are multivariate normal.
Linear	It is a linear combination of the present and past terms from a strict white noise process.

Gaussian process

A process is called Gaussian if the multivariate distribution of the consecutive variables $(X_{t+1},X_{t+2},...,X_{t+k})$ is multivariate normal for all integers t and k. A stationary Gaussian process is always strictly stationary, because then the first- and second-order moments completely determine the multivariate distributions.

Why white noise?

Autocovariances $$\lambda_\tau = \operatorname{cov}(X_t, X_{t+\tau}) = E[(X_t - \mu)(X_{t+\tau} - \mu)]$$

Autocorrelation of a stationary process (the variance is constant)

$\rho_0 = 1$, $\rho_\tau = \rho_{-\tau}$

$$\rho_\tau = \operatorname{cov}(X_t, X_{t+\tau})/\lambda_0 = \lambda_\tau/\lambda_0$$

Spectral density function

$$s(\omega) = \frac{\lambda_0}{2\pi}\left[1 + 2\sum_{\tau=1}^{\infty}\rho_\tau \cos(\tau\omega)\right]$$

The integral of $s(\omega)$ from 0 to 2π equals λ_0. High values of $s(\omega)$ might indicate cyclical behavior with the period of one cycle equal to $2\pi/\omega$ time units. For a white noise the spectral density function is the same constant for all frequencies ω

Definition B.15 *Let* $\{X_t;\ t \geq 0\}$ *be a random process, defined on the probability space* $(\Omega, \mathcal{F}, P)$. *We then define the* **sigma-algebra generated by X over the interval** $[0, t]$ *by*

$$\mathcal{F}_t^X = \sigma\{X_s;\ s \leq t\}.$$

The intuitive interpretation is that "$\mathcal{F}_t^X$ is the information generated by observing X over the time interval $[0, t]$". There is in general no very explicit description of $\mathcal{F}_t^X$, but it is not hard to show that $\mathcal{F}_t^X$ is generated by all events of the form $\{X_s \in B\}$ for all $s \leq t$ and all Borel sets B.

Definition B.16 *A* **filtration** $\mathcal{F} = \{\mathcal{F}_t\}_{t \geq 0}$ *on the probability space* $(\Omega, \mathcal{F}, P)$ *is an indexed family of sigma-algebras on* Ω *such that*

Björk : "Arbitrage Theory in Continuous Time"

- $$\mathcal{F}_t \subseteq \mathcal{F}, \quad \forall t \geq 0.$$
- $$s \leq t \Rightarrow \mathcal{F}_s^X \subseteq \mathcal{F}_t^X.$$

Given a filtration $\mathcal{F}$ *as above, the sigma-algebra* $\mathcal{F}_\infty$ *is defined as*

$$\mathcal{F}_\infty = \bigvee_{t \geq 0} \mathcal{F}_t.$$

- *Two sigma-algebras $\mathcal{G}, \mathcal{H} \subseteq \mathcal{F}$ are independent if*

$$P(G \cap H) = P(G) \cdot P(H),$$

for all $G \in \mathcal{G}$ and all $H \in \mathcal{H}$.

Björk : "Arbitrage Theory in Continuous Time"

- *Two random variables X and Y are independent if the sigma-algebras $\sigma\{X\}$ and $\sigma\{Y\}$ are independent.*
- *Two stochastic processes X and Y are independent if the sigma-algebras $\sigma\{X_t;\ t \geq 0\}$ and $\sigma\{Y_t;\ t \geq 0\}$ are independent.*
- *An indexed family $\{\mathcal{G}_\gamma;\ \gamma \in \Gamma\}$ of sigma-algebras, where $\mathcal{G}_\gamma \in \mathcal{F}$ for each $\gamma \in \Gamma$ are mutually independent if*

$$P\left(\bigcap_{i=1}^{n} G_n\right) = \prod_{i=1}^{n} P(G_n),$$

for every finite sub collection $G_1, \ldots, G_n$ where $G_i \in \mathcal{G}_{\gamma_i}$ and where $\gamma_i \neq \gamma_j$ for $i \neq j$. The extension to random variables and processes is the obvious one.

Proposition B.21 *Suppose that the random variables X and Y are independent. Assume furthermore that X, Y, and XY are in L^1. Then we have*

$$E[X \cdot Y] = E[X] \cdot E[Y]. \tag{B.1}$$

Consider a fixed probability space $(\Omega, \mathcal{F}, P)$, and suppose that A and B are events in $\mathcal{F}$ with $P(B) \neq 0$. We recall the elementary definition of conditional probability.

Definition B.23 *The probability of A, conditional on B is defined by*

$$P(A|B) = \frac{P(A \cap B)}{P(B)}. \tag{B.2}$$

The intuition behind this definition is as follows:

- The probability for any event A is the fraction of the total mass which is located on A, so
$$P(A) = \frac{P(A)}{P(\Omega)}$$
- When we condition on B, we **know** that B has happened. Thus the effective sample space is now B rather than Ω. This explains the normalizing factor in the nominator of (B.2).
- The only part of A that can occur if we know that B has occurred is precisely $A \cap B$.

What we are looking for is now a sensible definition of the object

$$E[X|\mathcal{G}],$$

where X is a random variable and $\mathcal{G}$ is a sigma-algebra included in $\mathcal{F}$. The interpretation should be that $E[X|\mathcal{G}]$ is "the expectation of X given that we have access to the information in $\mathcal{G}$". It is not trivial to formalize this rather vague notion, so we start with some heuristics.

Björk : "Arbitrage Theory in Continuous Time"

Definition B.24 *Suppose $B \in \mathcal{F}$ with $P(B) > 0$, and that $X \in L^1(\Omega, \mathcal{F}, P)$. Then "the conditional expectation of X given B" is defined by*

$$E[X|B] = \frac{1}{P(B)} \int_B X(\omega)\, dP(\omega).$$

Definition B.25 *With assumptions as above, and also assuming that $P(A_n) > 0$ for all n, we define $E[X|\mathcal{P}]$, "the conditional expectation of X given the information in $\mathcal{P}$" by*

$$E[X|\mathcal{P}](\omega) = \sum_{n=1}^{K} ID_{A_n}(\omega) E[X|A_n], \tag{B.3}$$

i.e.

$$E[X|\mathcal{P}](\omega) = \frac{1}{P(A_n)} \int_{A_n} X(\omega) DPP(\omega), \quad \text{when } \omega \in A_n. \tag{B.4}$$

Björk : "Arbitrage Theory in Continuous Time"

We now consider a slightly more general case, where we are given a finite partition $\mathcal{P} = \{A_1, \ldots, A_k\}$. Having access to the information contained in $\mathcal{P}$ is, according to our earlier discussion, equivalent to knowing exactly in which of the components $\{A_1, \ldots, A_k\}$ the outcome lies.

With assumptions as above, and also assuming that $P(A_n) > 0$ for all n, we define $E[X|\mathcal{P}]$, "the conditional expectation of X given the information in $\mathcal{P}$" by

$$E[X|\mathcal{P}] = \sum I_{A_n} E[X|A_n]$$

We now consider a slightly more general case, where we are given a finite partition $\mathcal{P} = \{A_1, \ldots, A_k\}$. Having access to the information contained in $\mathcal{P}$ is, according to our earlier discussion, equivalent to knowing exactly in which of the components $\{A_1, \ldots, A_k\}$ the outcome lies.

With assumptions as above, and also assuming that $P(A_n) > 0$ for all n, we define $E[X|\mathcal{P}]$, the conditional expectation of X given the information in $\mathcal{P}$ by

$$E[X|\mathcal{P}] = \sum I_{A_n} E[X|A_n]$$

We now would like to extend this definition to the case when we condition on a general σ-algebra. This is not straightforward since we had to assume $P(A_n) > 0$

Let $(\Omega, \mathcal{F}, P)$ be a probability space and X a random variable in $L^1(\Omega, \mathcal{F}, P)$. Let $\mathcal{G}$ be a sub- σ-algebra of $\mathcal{F}$ and Z a random variable such that

(i) Z is $\mathcal{G}$-measurable

(ii) For every $G \in \mathcal{G}$ it holds that $\int_G Z(\omega)dP(\omega) = \int_G X(\omega)dP(\omega)$

Then we say that Z is the conditional expectation of X given the sigma-algebra $\mathcal{G}$. In that case we denote $Z = E[X|\mathcal{G}]$

The conditional expectation is the optimal conditional predictor

It is well known that $E[X]$ is the optimal mean square deterministic predictor of X. The corresponding result for conditional expectations is as follows.

Proposition: Let $(\Omega, \mathcal{F}, P)$ be a probability space, let $\mathcal{G}$ be a sub- σ-algebra of $\mathcal{F}$ and let X be a square integrable random variable.
Consider the problem of minimizing $E[(X-Z)^2]$ where Z is allowed to vary over the class of all square integrable $\mathcal{G}$-measurable random variables.
The optimal solution is then given by $Z= E[X|\mathcal{G}]$

Information flows and filtrations

Let X be any given stochastic process. In the sequel it will be important to define "the information generated by X" as time goes by.

The symbol $\mathcal{F}_t^X$ denotes the information generated by X on the interval [0, t] or alternatively "what has happened to X over the interval [0,t] ".

If, based upon observations of the trajectory $\{X(s);\ 0 \leq s \leq t\}$, it is possible to decide whether a given event A has occurred or not, then we write this as $A \in \mathcal{F}_t^X$ or say that "A is $\mathcal{F}_t^X$-measurable".

If the value of a given stochastic variable Z can be completely determined given observations of the trajectory $\{X(s);\ 0 \leq s \leq t\}$, then we also write $Z \in \mathcal{F}_t^X$

If Y is a stochastic process such that we have $Y_t \in \mathcal{F}_t^X$ for all $t \geq 0$ then we say that Y is adapted to the filtration $\mathcal{F}_t^X$

Let $(\Omega, \mathcal{F}, P, \underline{\mathcal{F}})$ be a filtered probability space, and let X be a random process in continuous or discrete time.

The process X is an $\underline{\mathcal{F}}$-martingale if

1. X is $\underline{\mathcal{F}}$-adapted.
2. $X_t \in L^1$ for each t.
3. For every $0 \leq s \leq t$ one has $X_s = E[X_t \mid \mathcal{F}_s]$

If the equality sign is replaced by $\leq$ ($\geq$) then X is said to be a submartingale (supermartingale).

Note that the martingale property is always with respect to some given filtration.

Martingale theory in discrete time is easy, martingale theory in continuous time is rather complicated : we typically want our processes to be codlag (continue `a droite limit a gauche) and the filtration E must have some regularity properties.

It follows immediately from the definition, that a martingale is characterized by the property that the conditional expectations of a forward increment equals zero, i.e. that $E[X_t - X_s \mid \mathcal{F}_s] = 0$ for all $s \leq t$

Examples:

1. Take Y any integrable random variable on the filtered space $(\Omega, \mathcal{F}, P, \underline{\mathcal{F}})$ and define the process X by $X_t = E[Y \mid \mathcal{F}_t]$, $t \geq 0$

On a compact interval $[0, T]$ any given martingale M is always generated by its final value M_T by the formula $M_t = E[M_T \mid \mathcal{F}_t]$

2. If X is a process with independent increments on $(\Omega, \mathcal{F}, P, \underline{\mathcal{F}})$ and if also $E[X_t - X_s] = 0$, for all s, t, then X is a martingale.

3. Let $\{Z_n, n = 1,2,. \ . \ .\}$ be a family of independent integrable random variables, and define the discrete time process X by $X_n = \sum_{i=1}^{n} Z_i$ then X is a martingale w.r.t. the filtration $\mathcal{F}^X$.

There is a close connection between martingale theory, the theory of convex functions, and the theory of harmonic functions. The correspondence is as follows:

Martingale theory	Convex theory	Harmonic theory
martingale	linear function	harmonic function
submartingale	convex function	subharmonic function
supermartingale	concave function	superharmonic function

EMH: formalization

Under the EMH the stock price p_t already incorporates all relevant information and the only reason for prices to change between time t and time t+1 is the arrival of news.

Forecast errors $\varepsilon_{t+1} = p_{t+1} - E_t\, p_{t+1}$ should therefore be zero on average and uncorrelated with any information Θ_t that was available at time t

(RE: rational expectation) $\qquad p_{t+1} = E_t\, p_{t+1} + \varepsilon_{t+1}$

The forecast error is expected to be zero on average because prices only change on the arrival of 'news' which itself is a random variable.

The statement that the forecast error ε_{t+1} must be independent on Θ_t is known as the orthogonality property. If the forecasting error is serially correlated then the orthogonality property is violated: e.g. if ε_t follows a first-order autoregressive process AR(1) $\varepsilon_{t+1} = \rho\, \varepsilon_t + \nu_t$, where ν_t is a (white noise) random element (and by assumption is independent on the information Θ_t at time t). The forecast error $\varepsilon_t = p_t - E_{t-1}\, p_t$, is known at time t thus forms part of Θ_t. Therefore this period's forecast error ε_t has a predictable effect on next period's forecast error ε_{t+1} and using (RE) can be used to forecast future prices, thus violating EMH.

Diffusions and SDEs

Loosely speaking we say that a stochastic process X is a diffusion if its local dynamics can be approximated by a stochastic difference equation of the following type:

$X(t + \Delta t) - X(t) = \mu(t, X(t))\,\Delta t + \sigma(t, X(t))\,Z(t).$

Here Z(t) is a normally distributed disturbance term which is independent of everything which has happened up to time t, while μ and σ are given deterministic functions. The intuitive content is that, over the time interval $[t, t + \Delta t]$, the X-process is driven by two separate terms.

- A locally deterministic velocity $\mu(t, X(t))$ (local drift term)
- A Gaussian disturbance term, amplified by the factor $\sigma(t, X(t))$ (diffusion term)

Wiener process

A stochastic process W is called a Wiener process if the following conditions hold:

1. $W(0) = 0$.
2. The process W has independent increments, i.e. if $r < s \leq t < u$ then $W(u) - W(t)$ and $W(s) - W(r)$ are independent stochastic variables.
3. For $s < t$ the stochastic variable $W(t) - W(s)$ has the Gaussian distribution $N[0, (t-s)^{1/2}]$
4. W has continuous trajectori

Stochastic integrals

With probability 1 a Wiener trajectory is nowhere differentiable

$dX(t) = \mu(t, X(t))\,dt + \sigma(t, X(t))\,dW(t)$,

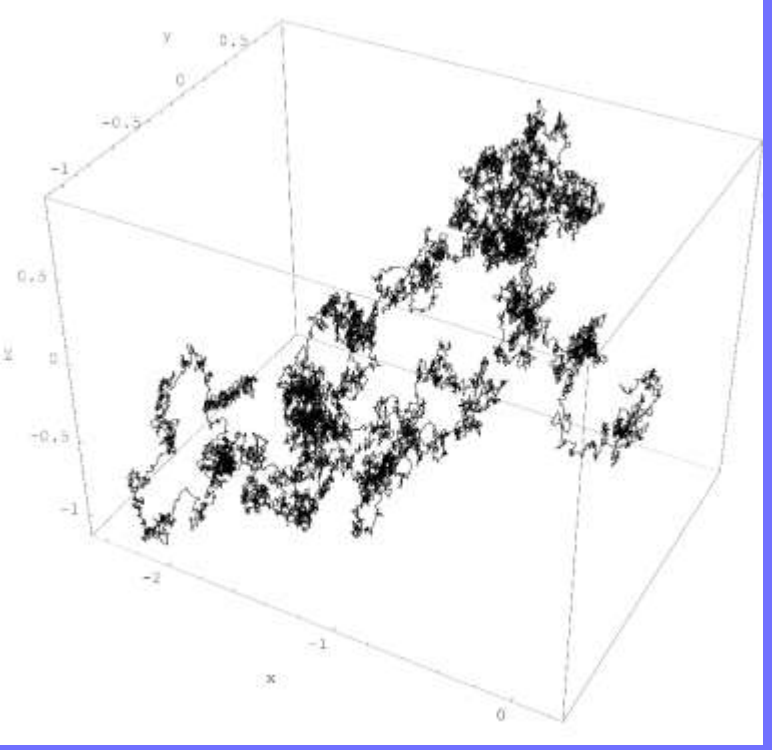
$X(0) = a$,

and it is now natural to interpret it as

$X(t) = a + \int_0^t \mu(s, X(s))\,ds + \int_0^t \sigma(s, X(s))\,dW(s)$

The natural interpretation of the dW-integral would be to view it as a Riemann-Stieltjes integral for each W-trajectory, but unfortunately this is not possible since one can show that the W-trajectories are of locally unbounded variation. Thus the stochastic dW-integral cannot be defined in a naive way.

If, however, we relax our demand that the dW-integral should be defined trajectorywise we can still proceed. It is in fact possible to give a global (L^2-)definition of these integrals : this new integral concept is called Itô integral



Approx. 60 b.c.: Lucretius in De rerum natura describes Brownian motion of dust particles in the air, observed “when sunbeams are admitted into a building and shed light on its shadowy places.”

1827: Robert Brown (a botanist) observes very irregular trajectories of fine particles suspendend in a liquid or gas

1900: Louis Bachelier → Brownian motion as a model of price fluctuations on the Paris stockmarket.

1905: Albert Einstein links Brownian motion to diffusion. Einstein recognized that if the predictions of statistical mechanics were correct, then *any* particle immersed in a "bath" of atoms must basically behave like a very large atom because it would be in thermodynamic equilibrium with the atoms in the bath.

By energy equipartition the particle's kinetic energy depends on temperature: for each degree of freedom the average kinetic energy is $k_B T/2$, where k_B is Boltzmann's constant and T is the temperature of the bath.

For spherical particles of radius r, if the liquid has a coefficient of viscosity k, then the diffusion coefficient is $D=RT/(6N\pi kr)$.
The distribution density function of the suspendend particles is
$f(x,t)=n \exp[-x^2/(4Dt)]/(4\pi Dt)$,
thus the mean value of the displacement of particles at time t is $\lambda=(2Dt)^{1/2}$

At room temperature, for particles with $r = 10^{-6}$ m suspendend in water the mean displacement in one minute is approximately $6\cdot 10^{-6}$ m.

This will be later confirmed by Perrin who observed the diffusion experimentally and used Einstein's formula for the diffusion coefficients to compute Avogadro's number N . The value thus calculated agreed excellently with other values obtained by entirely different methods in connection with other phenomena. In this way the discontinuity of matter was proved by him beyond doubt: an achievement rewarded with the 1926 Nobel Prize.

Brownian motion

1923: Norbert Wiener gives a rigorous construction of the corresponding measure P on the set Ω of all continuous functions on the interval [0,T] . The increments $S_t - S_s$, considered as random variables on the probability space (Ω,P), are Gaussian with means and variances proportional to the length $t - s$ of the time interval, and that they are independent for disjoint intervals. If this construction is carried out on a logarithmic scale, one obtains geometric Brownian motion, by now a standard model for the price fluctuation of a liquid financial asset, which was proposed by P.A. Samuelson in the 1960s.

Uncorrelated processes

The simplest possible autocorrelation occurs when a process is a collection of uncorrelated random variables so $\rho_0 = 1$, $\rho_\tau = 0$ for all $\tau > 0$

For an uncorrelated process the optimal forecast of the variable is simply the unconditional mean.

Uncorrelated processes are often used to model asset returns because they have some empirical support and they are coherent with the efficient markets hypothesis

Taylor, Asset Price Dynamics, Volatility and Prediction, P.U.P. (2005)

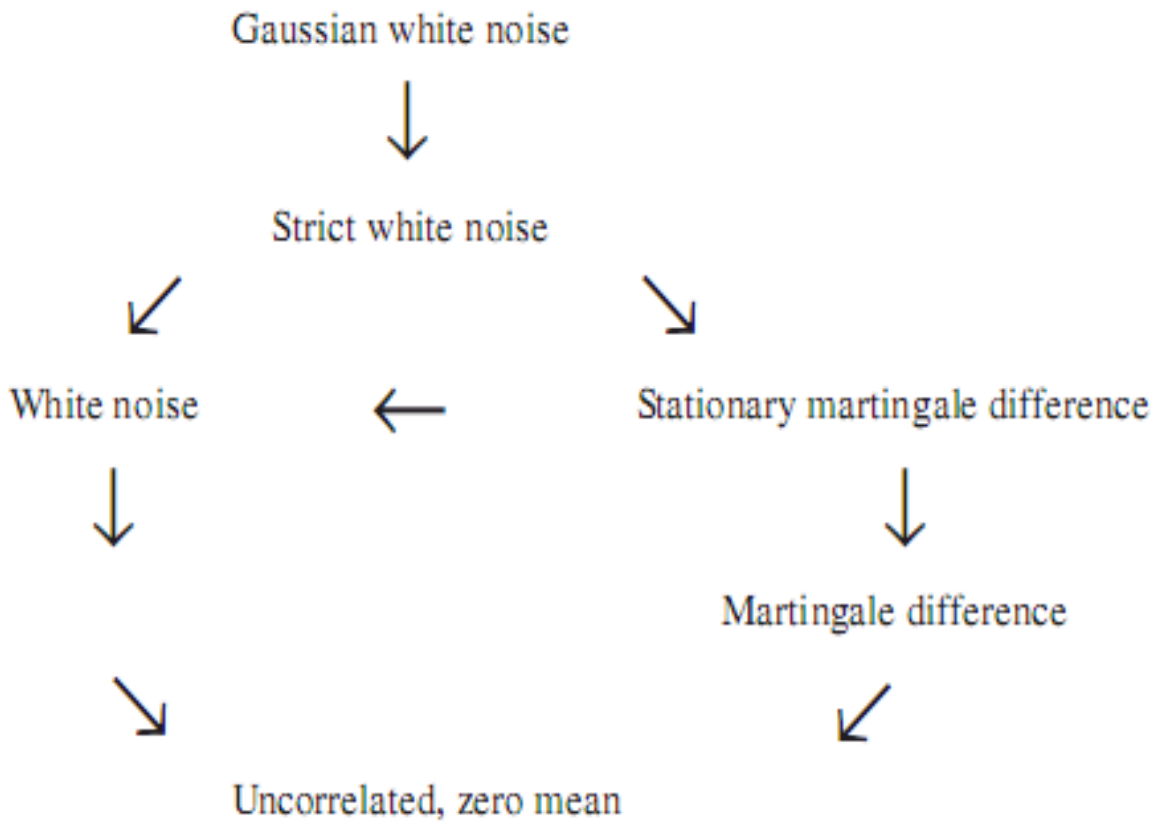


Figure 3.1. Relationships between categories of uncorrelated processes. An arrow pointing from one category to another indicates that all processes in the former category are also in the latter category *and* the converse is false: some processes in the latter category are not members of the former category. It is assumed that all processes have finite means and variances.

Random walk hypothesis (RWH)

1st possibility: returns have independent and identical distributions (i.i.d.). The i.i.d. hypothesis is not very relevant if we are interested in the predictability of returns. It will be rejected by an appropriate test if the conditional variances of returns have sufficient variation through time, but this may tell us nothing about the predictability of returns. For example, the statistically significant autocorrelation in absolute and squared returns rejects the i.i.d. hypothesis but it does not prove that returns can be predicted. Even if we test and reject the i.i.d. hypothesis using the autocorrelations of returns, we still cannot reject the hypothesis that returns are uncorrelated at the same significance level.

2nd possibility (quite general): replace identical distributions by identical means and independent distributions by uncorrelated distributions:

$E[r_t] = E[r_{t+\tau}]$ and $\mathrm{cov}(r_t, r_{t+\tau}) = 0$ for all t and all $\tau > 0$ (RWH1)

One does not even requires stationarity.

Linear predictors of r_{t+1} are defined by $f_{t+1} = \alpha + \sum_{i=0}^{\infty} \beta_i r_{t-i}$.
When RWH1 is true, the returns process is uncorrelated and hence the best linear prediction of a future return is its unconditional mean, which RWH1 assumes is a constant: $E[(r_{t+1} - f_{t+1})^2]$ is minimized by the constant predictor given by setting $\alpha = E[r_{t+1}]$ and all $\beta_i = 0$.

The definition RWH1 does not exclude the possibility that a nonlinear predictor is more accurate than the unconditional expectation. The unconditional mean is the best prediction when Samuelson's definition of RWH applies, namely,
$E[r_{t+1} \mid I_t] = \mu$ for some constant μ and for all times t and all return histories $I_t = \{r_{t-i}, i \geq 0\}$ (RWH2)

These conditions are the same as saying that returns have a stationary mean μ and that the process of excess returns, $\{r_t - \mu\}$, is a martingale difference.

RWH2 $\longrightarrow$ RWH1, whenever returns have finite variance. Most tests of the random walk hypothesis employ sample autocorrelations and are hence tests of RWH1. These tests reject RWH2 whenever they reject RWH1, as we assume returns have finite variance.
A stationary mean for returns appears in the definitions to ensure that the sample autocorrelations are consistent estimates. Asset pricing models do not, of course, require expected returns to be constant through time.

Random walks and market efficiency

Tests of the random walk hypothesis can provide insight into issues of market efficiency. Nevertheless, random walk tests should not be considered to be tests of the weak-form efficient market hypothesis (EMH).

First, consider the situation when the RWH is false. The EMH can then be true, for some definitions of market efficiency, or it too may be false. Prices can fully reflect the information in past prices, and thus the EMH holds, as defined by Fama (1976, 1991), when the RWH is false.

For example, conditional expected returns, $E[r_{t+1} \mid I_t]$, could depend on previous returns because the asset's risk premium follows a stationary, autocorrelated process. Or $E[r_{t+1} \mid I_t]$ could be a function of the conditional variance, $var(r_{t+1} \mid I_t)$.

Random walks and market efficiency

Another possibility is that some linear predictor is more accurate than prediction using a constant value but transaction costs exceed gross, risk-adjusted payoffs from trading. Then the EMH holds, as defined by Jensen (1978), yet the RWH is false. For example, returns could follow an MA(1) process with the moving-average parameter so close to zero that net trading profits are impossible. Efficiency might, however, be defined as a fair game for excess returns (LeRoy 1989) and then the EMH will be false whenever expected returns are constant and RWH2 is false.

Second, consider the situation when RWH1 is true. Then there may exist a nonlinear predictor which is more accurate than prediction using a constant value and, consequently, (i) RWH2 is false, (ii) the EMH can be false using the LeRoy definition, (iii) the EMH can be false for the Jensen definition when trading costs are sufficiently low, and (iv) the EMH can be false for Fama's definition as Jensen inefficiency implies Fama inefficiency. The existence of a successful nonlinear predictorwhenRWH1 is true is, however, a theoretical possibility which is unlikely to have practical relevance.

Variance-ratio tests (Lo-MacKinlay)

The variance of a multi-period return is the sum of single-period variances when the RWH is true. Several tests seek to exploit any divergence from this prediction, the most important being the variance-ratio test of Lo and MacKinlay (1988). To provide some intuition for the test, initially suppose that the stochastic process generating returns is stationary, with $V(1) = \mathrm{var}(r_t)$.

Two-period returns are the sum of two consecutive returns and their variance equals

$V(2) = \mathrm{var}(r_t + r_{t+1}) = \mathrm{var}(r_t) + \mathrm{var}(r_{t+1}) + 2\mathrm{cov}(r_t, r_{t+1}) = (2+2\rho_1)V(1)$,

with ρ_1 the first-lag autocorrelation of one-period returns. The two-period variance ratio is defined by

$$VR(2) = V(2)/(2V(1)) = 1 + \rho_1$$

The autocorrelation term is zero when the RWH applies and then the variance ratio is one. Otherwise, the RWH is false and the ratio can be either more or less than one.

Next consider N-period returns for any integer $N \geq 2$. When the RWH is true,

$V(N) = \mathrm{var}(r_t + r_{t+1} + \cdots + r_{t+N-1}) = \mathrm{var}(r_t) + \mathrm{var}(r_{t+1}) + \cdots + \mathrm{var}(r_{t+N-1}) = NV(1)$

and thus the variance ratio is unity for all N:

$VR(N) = V(N)/(NV(1)) = 1$

Variance-ratio tests (Lo-MacKinlay)

When the RWH is false, V(N) equals NV(1) plus the covariance terms between all pairs of distinct returns; thus

$V(N) = NV(1) + 2\sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \mathrm{cov}(r_{t+i-1}, r_{t+j-1}) = V(1)[N+2 \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \rho_{j-i}]$

The double summation can be simplified to give the variance ratio as

$VR(N) = 1 + 2N^{-1} \sum_{\tau=1}^{N-1} (N-\tau)\rho_{\tau}$

The empirical test uses observed returns to decide if a sample estimate of the variance ratio is compatible with the theoretical prediction. The test is most likely to reject the RWH when the ratio is far from one. This happens when a linear function of the first N − 1autocorrelations, namely

$(N-1)\rho_1 + (N-2)\rho_2 + \cdots + 2\rho_{N-2} + \rho_{N-1}$

is far from zero. The multiplier is $N - \tau$ for ρ_{τ} .All the multipliers are positive and they decrease as the lag increases. A variance-ratio test is therefore particularly appropriate when the alternative to randomness involves autocorrelations that all have the same sign and that decrease as the lag increases.

E.g.: autoregressive models with mean reversion in prices or with trends in prices

5.3.2 The Test Statistic

The researcher must first choose a value for N. Indeed, the choice can appear to be arbitrary. Suppose a set of n observed returns has average $\bar{r}$ and variance $\hat{V}(1) = \sum(r_t - \bar{r})^2/(n-1)$. An appropriate estimate of $V(N)$ is

$$\hat{V}(N) = \frac{n}{(n-N)(n-N+1)} \sum_{t=1}^{n-N+1} (r_t + r_{t+1} + \cdots + r_{t+N-1} - N\bar{r})^2 \quad (5.6)$$

and then the sample variance ratio is

$$\widehat{\mathrm{VR}}(N) = \frac{\hat{V}(N)}{N\hat{V}(1)}. \quad (5.7)$$

An appropriate estimate of $n\,\mathrm{var}(\hat{\rho}_\tau)$ is provided by

$$b_\tau = \frac{n\sum_{t=1}^{n-\tau} s_t s_{t+\tau}}{(\sum_{t=1}^{n} s_t)^2}, \quad \text{with } s_t = (r_t - \bar{r})^2, \quad (5.9)$$

and then an estimate of $n\,\mathrm{var}(\widehat{\mathrm{VR}}(N))$ is given by

$$v_N = \frac{4}{N^2} \sum_{\tau=1}^{N-1} (N-\tau)^2 b_\tau. \quad (5.10)$$

The above estimates are consistent when the RWH is true; they then converge to the parameters that they estimate as n increases. Finally, the standardized distribution of the sample variance ratio,

$$z_N = \frac{\widehat{\mathrm{VR}}(N) - 1}{\sqrt{v_N/n}}, \quad (5.12)$$

is approximately the standard normal distribution when the RWH is true. This is an asymptotic result, so the approximation becomes perfect as $n \to \infty$.

Taylor, Asset Price Dynamics, Volatility and Prediction, P.U.P. (2005)

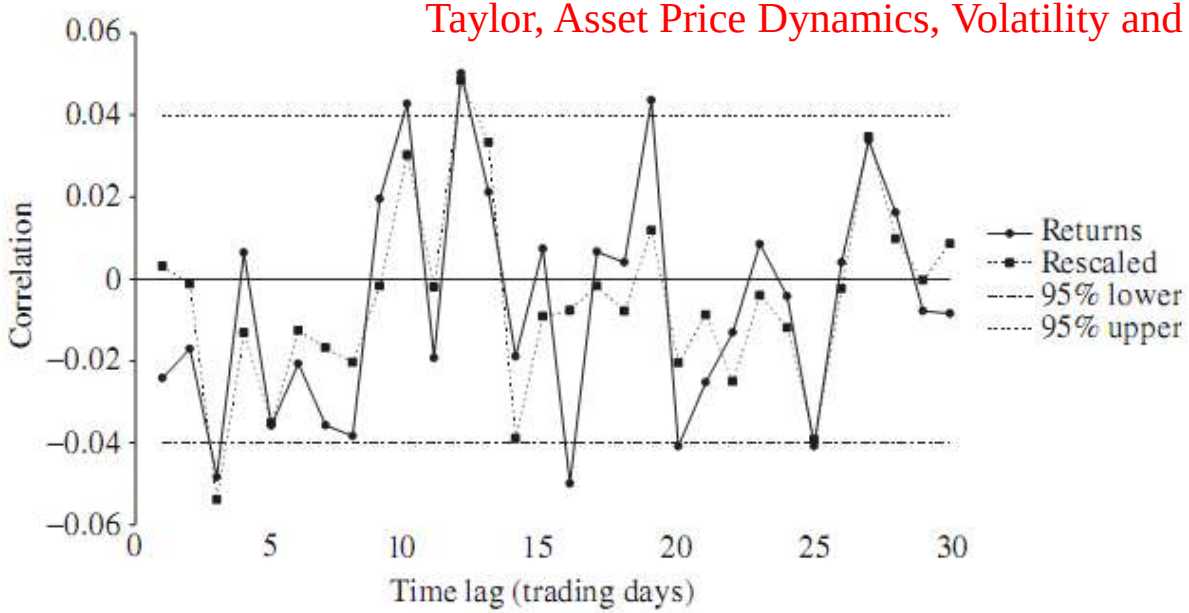


Figure 5.1. S&P 100 autocorrelations.

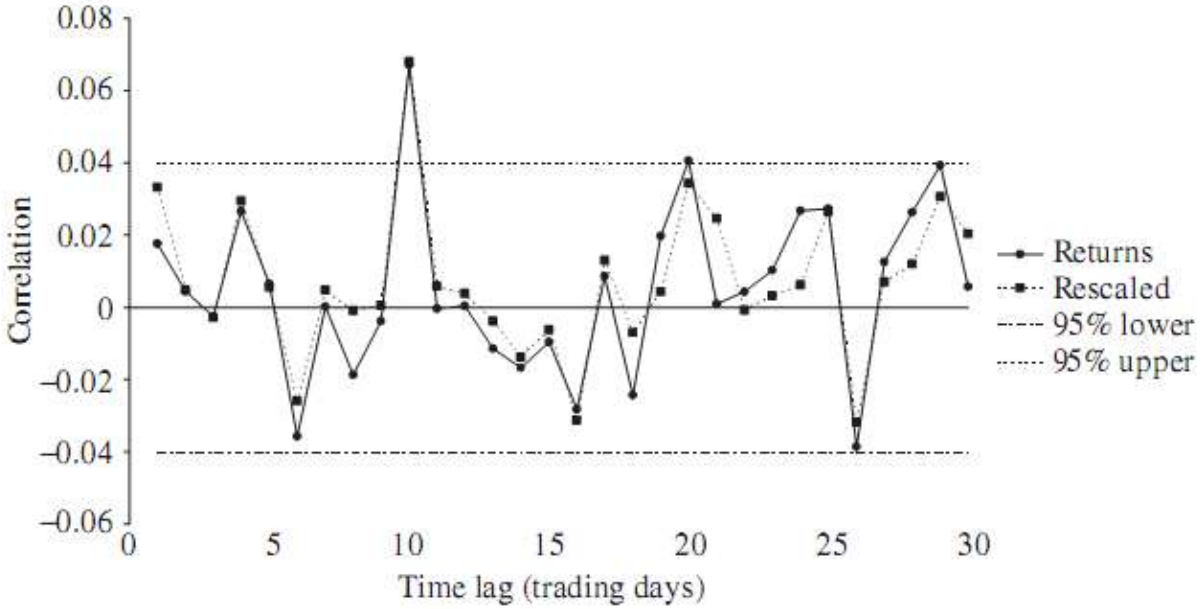


Figure 5.2. Spot DM/$ autocorrelations.

5.7 Random Walk Tests Using Rescaled Returns

Returns do not have constant conditional variances and this is the primary reason for their autocorrelations having more variability than those calculated from i.i.d. processes. The excess variability can often be reduced substantially if we can find a way to rescale returns that ensures the rescaled quantities have approximately constant conditional variances.

5.7.1 Definition

Rescaled returns are defined by

$$r_t^* = \frac{r_t - \bar{r}}{\sqrt{h_t}} \tag{5.20}$$

with h_t a conditional variance for period t calculated from returns observed until period $t-1$. When the RWH is true for the process generating returns we may also expect the hypothesis to be true for the process generating rescaled returns. When the RWH is false, the autocorrelations of the random variables that generate returns and rescaled returns can differ by important amounts. Reasons for this and

Test statistic		Returns z_2	Returns z_5	Returns z_{20}	Rescaled returns z_2	Rescaled returns z_5	Rescaled returns z_{20}
Series from 1991 to 2000							
S&P 100-share	S	−0.73	−1.41	−1.76	0.17	−0.92	−1.58
DM/$	S	0.73	0.80	0.32	1.59	1.55	1.41
Twenty further series							
S&P 500-share	S	4.00*	2.66*	0.62	5.20*	4.46*	1.90
S&P 500-share	F	−1.40	−1.43	−1.75	−0.75	−0.68	−1.02
Coca Cola	S	−1.24	−2.33*	−2.05*	0.16	−0.85	−1.06
General Electric	S	−0.92	−1.93	−1.27	−0.73	−0.82	−0.83
General Motors	S	0.57	−1.29	−0.75	1.34	−0.20	0.16
FT 100-share	S	2.51*	1.50	1.68	3.80*	3.57*	4.30*
FT 100-share	F	−0.47	−1.23	−0.51	0.72	0.25	2.13*
Glaxo	S	3.56*	1.85	0.48	5.88*	4.13*	2.24*
Marks & Spencer	S	1.96*	0.40	−1.44	2.80*	1.54	−0.22
Shell	S	2.34*	2.52*	0.34	4.68*	4.10*	1.13
Nikkei 225-share	S	1.83	−0.01	0.46	3.57*	2.69*	3.76*
Treasury bonds	F	0.71	0.47	0.48	1.73	0.91	1.46
3-month sterling bills	F	1.21	−0.30	0.40	4.91*	4.54*	4.70*
DM/$	F	−0.01	0.04	1.09	0.78	1.48	3.19*
Sterling/$	F	1.14	0.23	0.46	1.69	0.71	1.96
Swiss franc/$	F	−0.55	−0.57	0.49	−0.09	−0.10	1.43
Yen/$	F	−0.01	0.55	2.60*	0.29	0.94	3.36*
Gold	F	−1.88	−0.35	−0.47	−0.82	0.13	0.28
Corn	F	2.94*	1.82	2.33*	4.70*	3.14*	4.22*
Live cattle	F	0.56	0.94	−0.00	0.52	1.08	0.24

The crash week, commencing on 19 October 1987, is excluded from the time series. Stars identify test values that reject the RWH at the 5% level, for two-tailed tests. The test statistics are defined by equation (5.12).

Taylor, Asset Price Dynamics, Volatility and Prediction, P.U.P. (2005)