Dynamics and time series: theory and applications

Stefano Marmi
Scuola Normale Superiore

Lecture 2, Jan 15, 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: A non-random walk down Wall Street. (Tue Mar 24, 2 pm – 4 pm Bianchi)

- 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 (TBA)

Dynamical systems

- A dynamical system is a couple (phase space, time evolution law)
- The phase space is the set of all possible states (i.e. initial conditions) of our system
- Each initial condition uniquely determines the time evolution (determinism)
- The system evolves in time according to a fixed law (iteration of a map, differential equation, etc.)
- Often (but not necessarily) the evolution law is not linear

Ergodic theory

- The focus of the analysis is mainly on the asymptotic ditribution of the orbits, and not on transient phenomena. Ergodic theory is an attempt to study the statistical behaviour of orbits of

  dynamical systems restricting the attention to their asymptotic distribution. One waits until all transients have been wiped off and looks for an invariant probability measure describing the distribution of typical orbits.

Measure theory vs. probability theory

Table 1.1. Comparison of terminology

Measure Theory	Probability Theory
a probability measure space X	a sample space Ω
$x \in X$	$\omega \in \Omega$
a σ-algebra $\mathcal{A}$	a σ-field $\mathcal{F}$
a measurable subset A	an event E
a probability measure μ	a probability P
$\mu(A)$	$P(E)$
a measurable function f	a random variable X
$f(x)$	x, a value of X
a characteristic function χ_E	an indicator function 1_E
Lebesgue integral $\int_X f \, d\mu$	expectation $E[X]$
almost everywhere	almost surely, or with probability 1
convergence in L^1	convergence in mean
convergence in measure	convergence in probability
conditional measure $\mu_A(B)$	conditional probability $\Pr(B\|A)$

Stochastic or chaotic?

- An important goal of time-series analysis is to determine, given a times series (e.g. HRV) if the underlying dynamics (the heart) is:
  - Intrinsically random
  - Generated by a deterministic nonlinear chaotic system which generates a random output
  - A mix of the two (stochastic perturbations of deterministic dynamics)

Randomness and the physical law

- It may well be that the universe itself is completely deterministic (though this depends on what the “true” laws of physics are, and also to some extent on certain ontological assumptions about reality), in which case randomness is simply a mathematical concept, modeled using such abstract mathematical objects as probability spaces. Nevertheless, the concept of *pseudorandomness*- objects which “behave” randomly in various statistical senses - still makes sense in a purely deterministic setting. A typical example are the digits of π=3.14159265358979323385…this is a deterministic sequence of digits, but is widely believed to behave pseudorandomly in various precise senses (e.g. each digit should asymptotically appear 10% of the time). If a deterministic system exhibits a sufficient amount of pseudorandomness, then random mathematical models (e.g. statistical mechanics) can yield accurate predictions of reality, even if the underlying physics of that reality has no randomness in it.

http://terrytao.wordpress.com/2007/04/05/simons-lecture-i-structure-and-randomness-in-fourier-analysis-and-number-theory/

Probability, statistics and the problem of induction

- The probability of an event (if it exists) is almost always impossible to be known a-priori
- The only possibility is to replace it with the frequencies measured by observing how often the event occurs
- The problem of backtesting
- The problem of ergodicity and of typical points: from a single series of observations I would like to be able to deduce the invariant probability
- Bertrand Russell’s chicken (turkey nella versione USA)

Bertrand Russel

(The Problems of Philosophy,
Home University Library, 1912. Chapter VI On Induction) Available at the page http://www.ditext.com/russell/rus6.html

Domestic animals expect food when they see the person who feeds them. We know that all these rather crude expectations of uniformity are liable to be misleading. The man who has fed the chicken every day throughout its life at last wrings its neck instead, showing that more refined views as to the uniformity of nature would have been useful to the chicken.

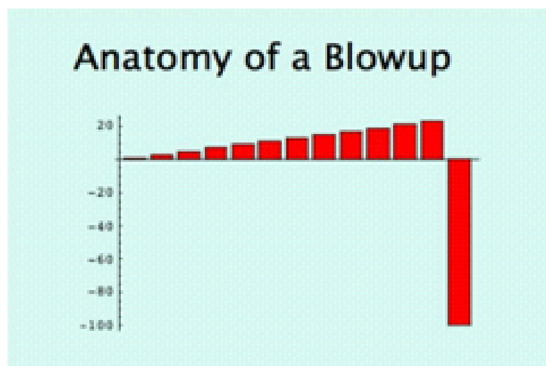


Figure 1 *My classical metaphor: A Turkey is fed for a 1000 days—every days confirms to its statistical department that the human race cares about its welfare "with increased statistical significance". On the 1001st day, the turkey has a surprise.*

http://www.edge.org/3rd_culture/taleb08/taleb08_index.html

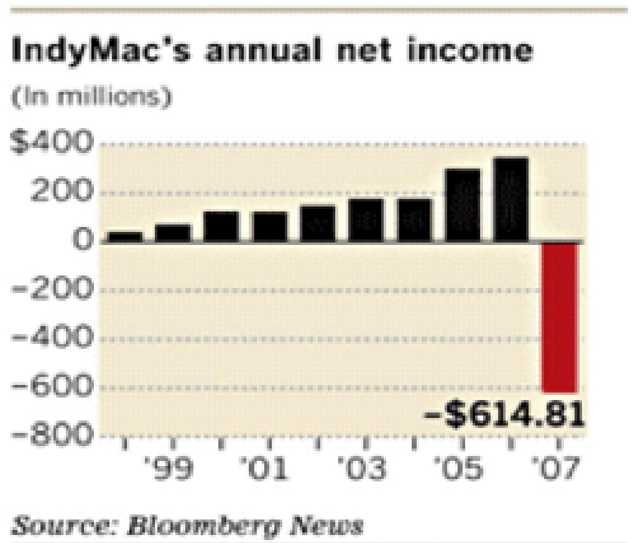


Figure 2 *The graph above shows the fate of close to 1000 financial institutions (includes busts such as FNMA, Bear Stearns, Northern Rock, Lehman Brothers, etc.). The banking system (betting AGAINST rare events) just lost > 1 Trillion dollars (so far) on a single error, more than was ever earned in the history*

http://www.edge.org/3rd_culture/taleb08/taleb08_index.html

Payoff from mildly OTM UK Sterling Short Option, 1988-2008

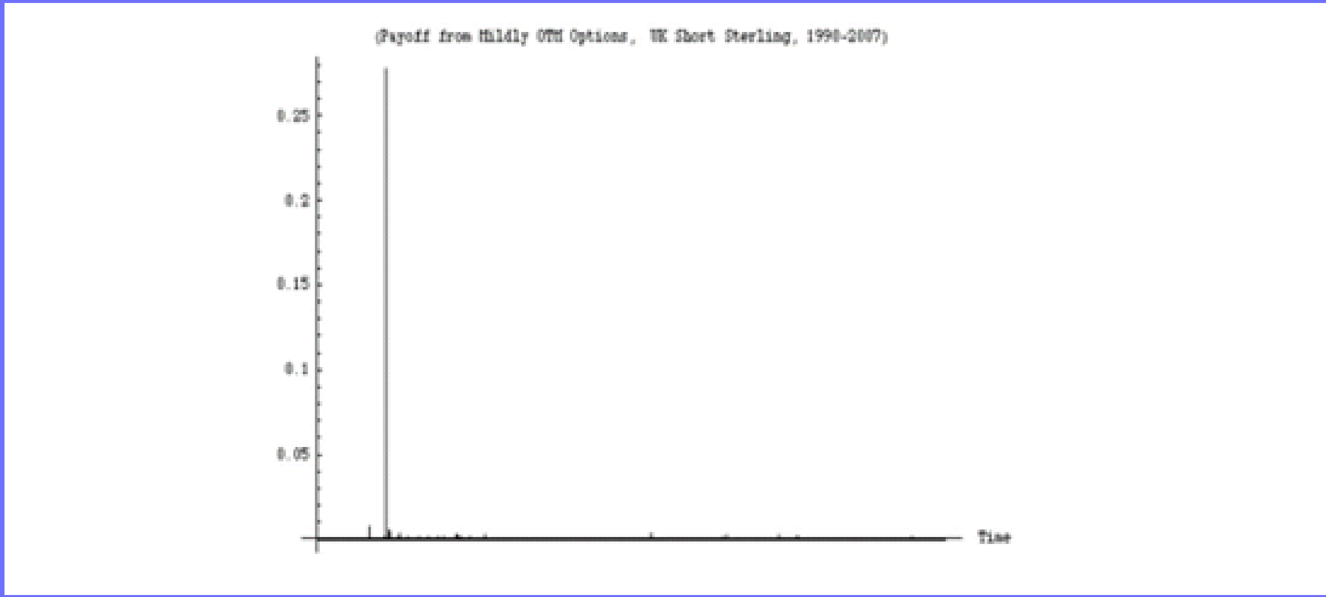


Figure 3 The graph shows the daily variations a derivatives portfolio exposed to U.K. interest rates between 1988 and 2008. Close to 99% of the variations, over the span of 20 years, will be represented in 1 single day—the day the European Monetary System collapsed. As I show in the appendix, this is typical with ANY socio-economic variable (commodity prices, currencies, inflation numbers, GDP, company performance, etc.). No known econometric statistical method can capture the probability of the event with any remotely acceptable accuracy (except, of course, in hindsight, and "on paper"). Also note that this applies to surges on electricity grids and all manner of modern-day phenomena.

http://www.edge.org/3rd_culture/taleb08/taleb08_index.html

Measure-preserving transformations

X phase space, μ probability measure

$\Phi: X \to \mathbf{R}$ observable (a measurable function, say L^2). Let A be subset of X (event).

$\mu(\Phi) = \int_X \Phi \, d\mu$ is the expectation of Φ

$T: X \to X$ induces a time evolution

on observables $\Phi \to \Phi \circ T$

on events $A \to T^{-1}(A)$

T is measure preserving if $\mu(\Phi) = \mu(\Phi \circ T)$ i.e. $\mu(A) = \mu(T^{-1}(A))$

Birkhoff theorem and ergodicity

Birkhoff theorem: if T preserves the measure μ then with probability one the time averages of the observables exist (statistical expectations). The system is ergodic if these time averages do not depend on the orbit (statistics and a-priori probability agree)

$$\frac{1}{N}\sum_{0}^{N-1}\varphi\circ T^{i}(x) := \frac{1}{N}S_N\varphi(x) \longrightarrow \int_X \varphi(t)\,d\mu(t)$$

$$\frac{1}{N}\#\left\{i\in[0,N), T^{i}(x)\in A\right\} \longrightarrow \mu(A)$$

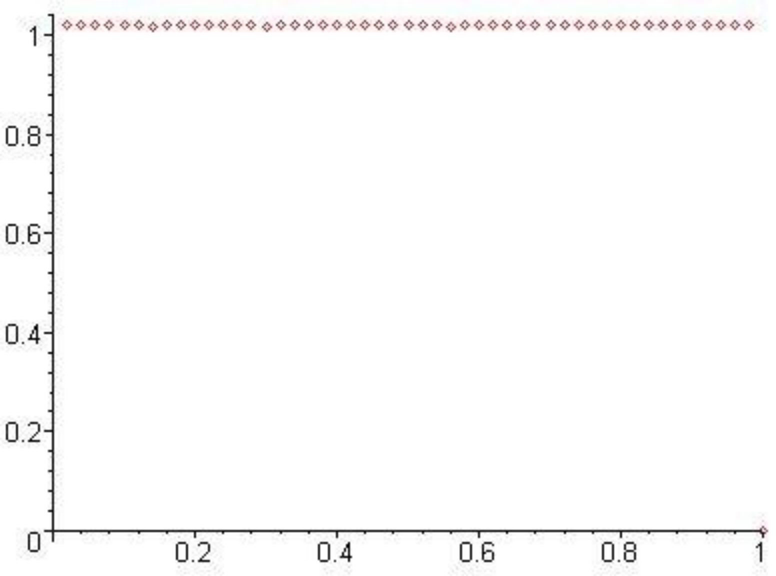
Law of large numbers: Statistics of orbits = a-priori probability

“Historia magistra vitae” or the mathematical foundation of backtesting

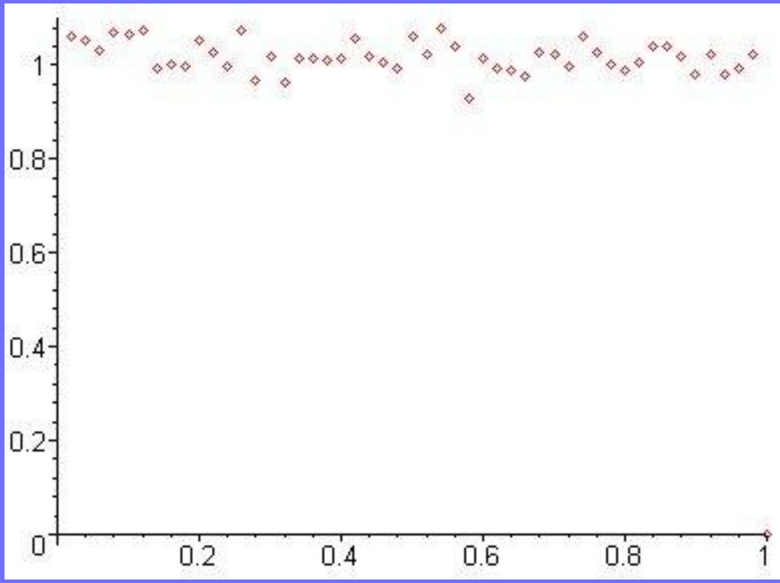
- Without assuming ergodicity, Birkhoff theorem shows that:

- Time averages exist and they give rise to an experimental statistics to compare with theory
- Past and future time averages agree almost everywhere

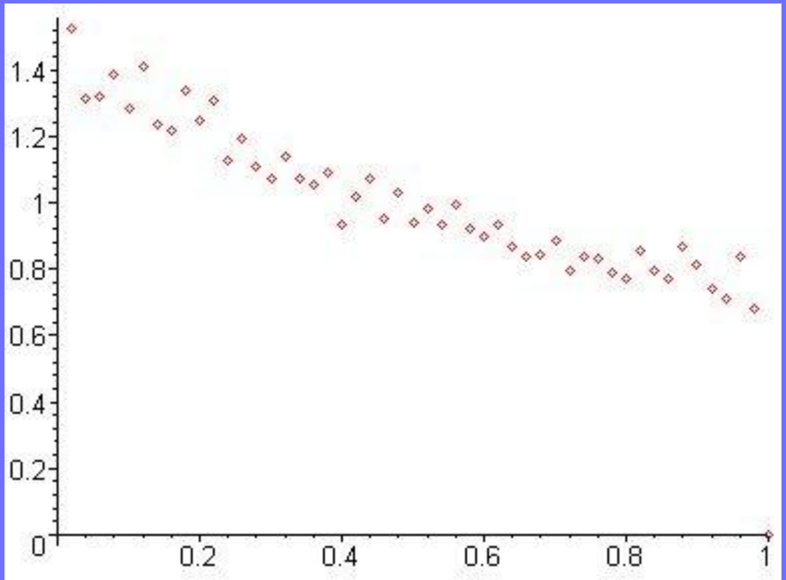
Statistical distribution of frequencies of vists



Rotation

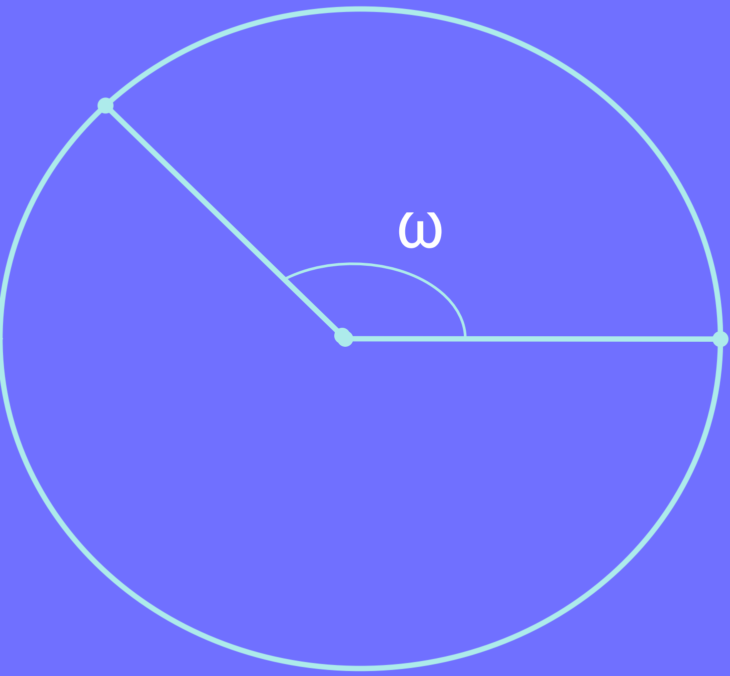


Doubling map



Gauss map

The simplest dynamical systems

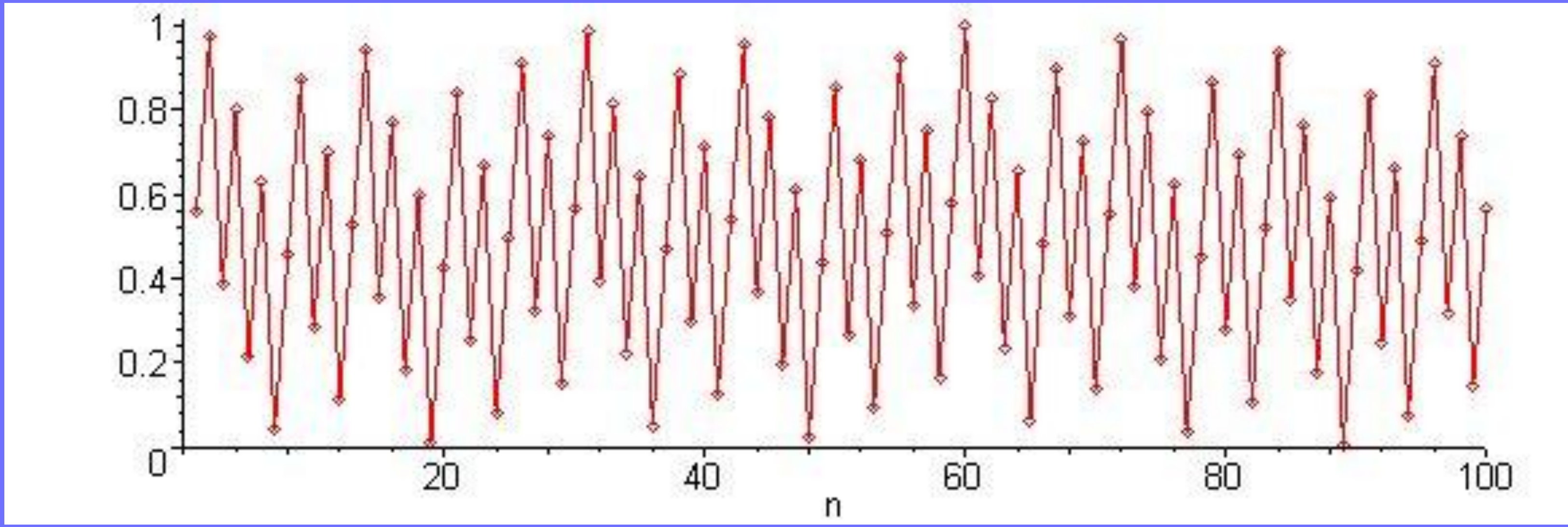


- The phase space is the circle: **S=R/Z**
- Case 1: quasiperiodic dynamics

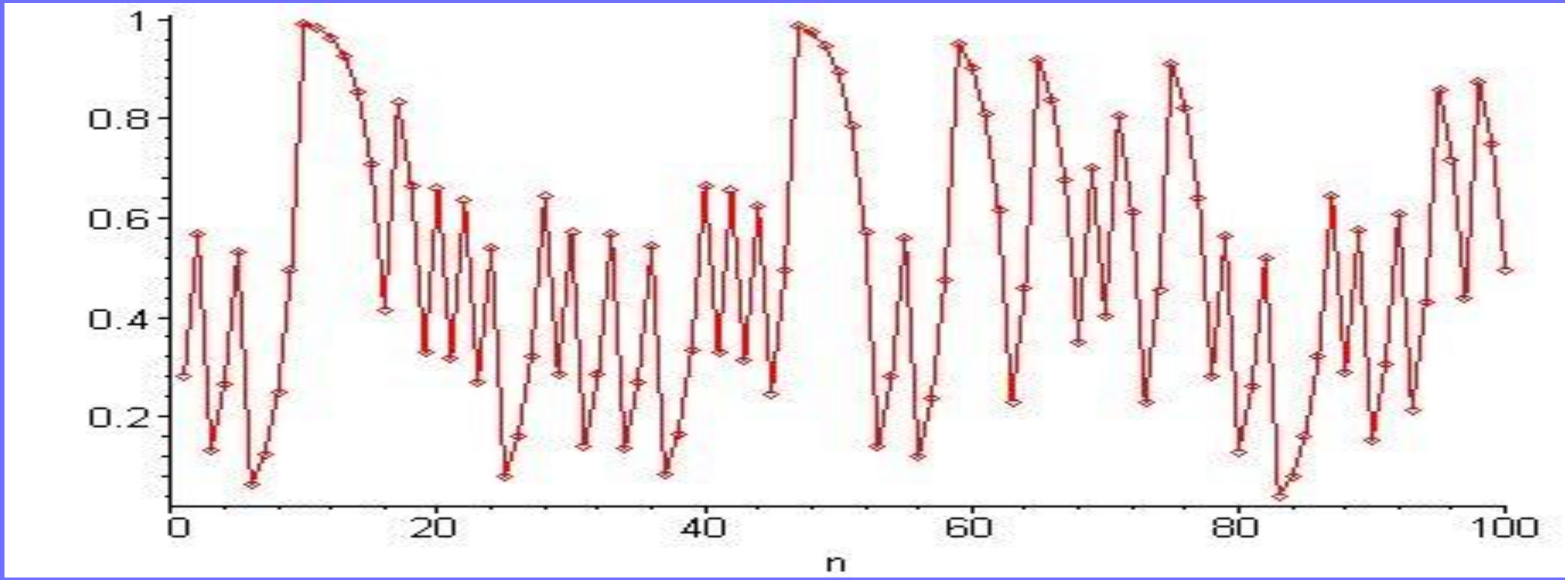
  $\theta(n+1)=\theta(n)+\omega \pmod 1$

  (ω irrational)
- Case 2: chaotic dynamics

  $\theta(n+1)=2\theta(n) \pmod 1$



1
0.8
0.6
0.4
0.2
0
20
40
60
80
100
n



1
0.8
0.6
0.4
0.2
0
20
40
60
80
100
n

Recurrence times

- A point is recurrent when it is a point of accumulation of its future (and past) orbit
- Poincarè recurrence: given a dynamical system T which preserves a probability measure μ and a set of positive measure E a point x of E is almost surely recurrent
- First return time of x in E:

  $$R(x,E)=\min\{n>0,\ T^n x \in E\}$$

- E could be an element of a partition of the phase space (symbolic dynamics): this point of view is very important in applications (e.g. the proof of optimality of the Lempel-Ziv data compression algorithm)

Kac's Lemma

- If T is ergodic and E has positive measure then

$$\int_E R(x,E)\,d\mu(x)=1 ,$$

i.e. $R(x,E)$ is of the order of $1/\mu(E)$: the average length of time that you need to wait to see a particular symbol is the reciprocal of the probability of a symbol. Thus, we are likely to see the high-probability strings within the window and encode these strings efficiently.

The ubiquity of "cycles" (as long as they last…)

Furstenberg's recurrence: If E is a set of positive measure in a measure-preserving system, and k is a positive integer, then there are infinitely many integers n for which

$$\mu(E \cap T^{-n}(E) \cap \ldots \cap T^{-(k-1)n}(E) > 0$$

- There are few persons, even among the calmest thinkers, who have not occasionally been startled into a vague yet thrilling half-credence in the supernatural, by *coincidences* of so seemingly marvellous a character that, as *mere* coincidences, the intellect has been unable to receive them. Such sentiments -- for the half-credences of which I speak have never the full force of *thought* -- such sentiments are seldom thoroughly stifled unless by reference to the doctrine of chance, or, as it is technically termed, the Calculus of Probabilities. Now this Calculus is, in its essence, purely mathematical; and thus we have the anomaly of the most rigidly exact in science applied to the shadow and spirituality of the most intangible in speculation. (Egdar Allan Poe, The mistery of Marie Roget)

ψ,φ observables with expectations $\mu(\psi)$ and $\mu(\varphi)$

$\sigma(\psi)^2 = [\,(\mu(\psi^2) - \mu(\psi)^2]$ **variance**

The **correlation coefficient** of ψ,φ is

$$\rho(\psi,\varphi) = \text{covariance}(\psi,\varphi) / (\sigma(\psi)\,\sigma(\varphi))$$
$$= \mu\,[(\psi - \mu(\psi))(\varphi - \mu\,(\varphi))] / (\sigma(\psi)\,\sigma(\varphi))$$
$$= \mu\,[\psi\,\varphi - \mu(\psi)\mu\,(\varphi)] / (\sigma(\psi)\,\sigma(\varphi))$$

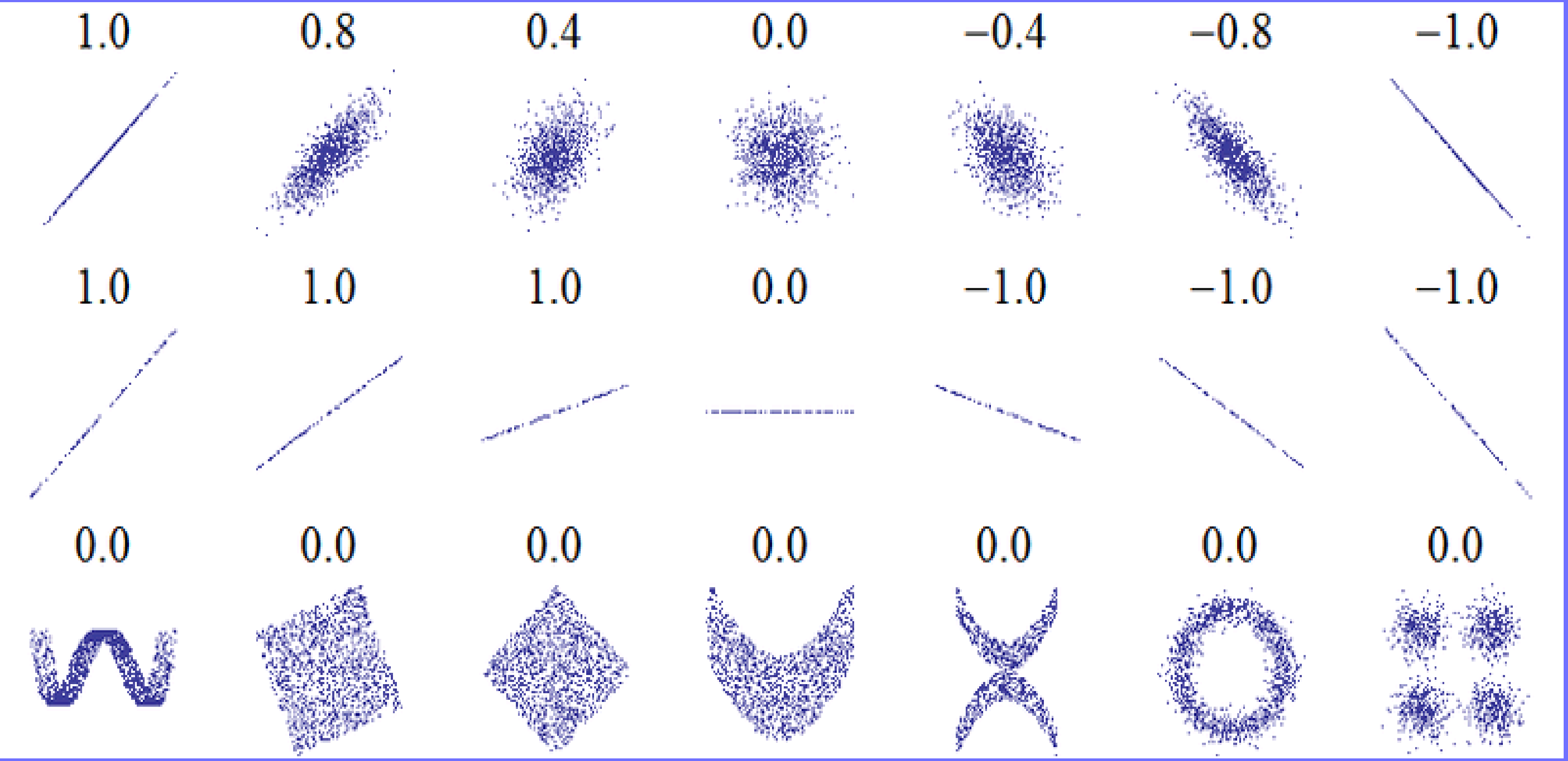
The correlation coefficient varies between -1 and 1 and equals 0 for independent variables but this is only a necessary condition (e.g. φ uniform on [-1,1] has zero correlation with its square)

If we have a series of n measurements of X and Y written as x(i) and y(i) where i = 1, 2, ..., n, then the Pearson product-moment correlation coefficient can be used to estimate the correlation of X and Y . The Pearson coefficient is also known as the "sample correlation coefficient". The Pearson correlation coefficient is then the best estimate of the correlation of X and Y . The Pearson correlation coefficient is written:

$$r_{xy} = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{(n-1)s_x s_y} = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{\sqrt{n\sum x_i^2 - (\sum x_i)^2}\ \sqrt{n\sum y_i^2 - (\sum y_i)^2}}.$$

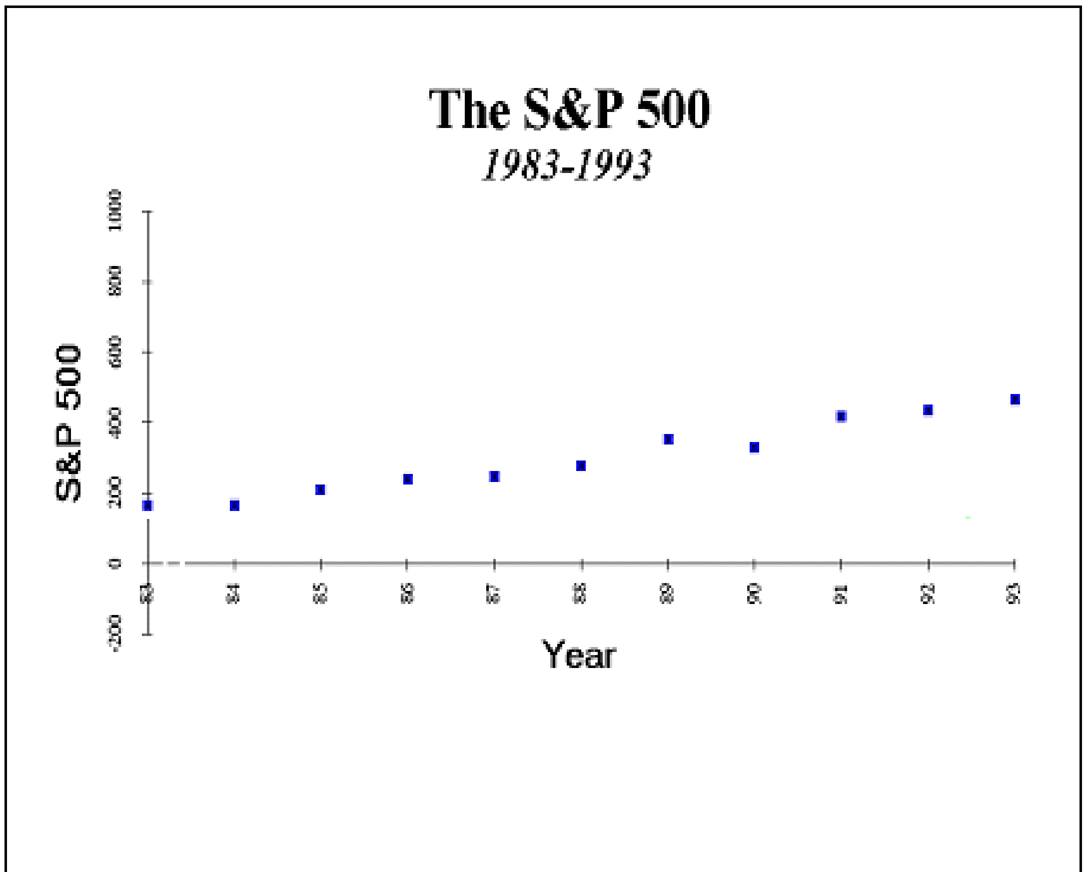
$$r_{xy} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{(n-1)s_x s_y},$$

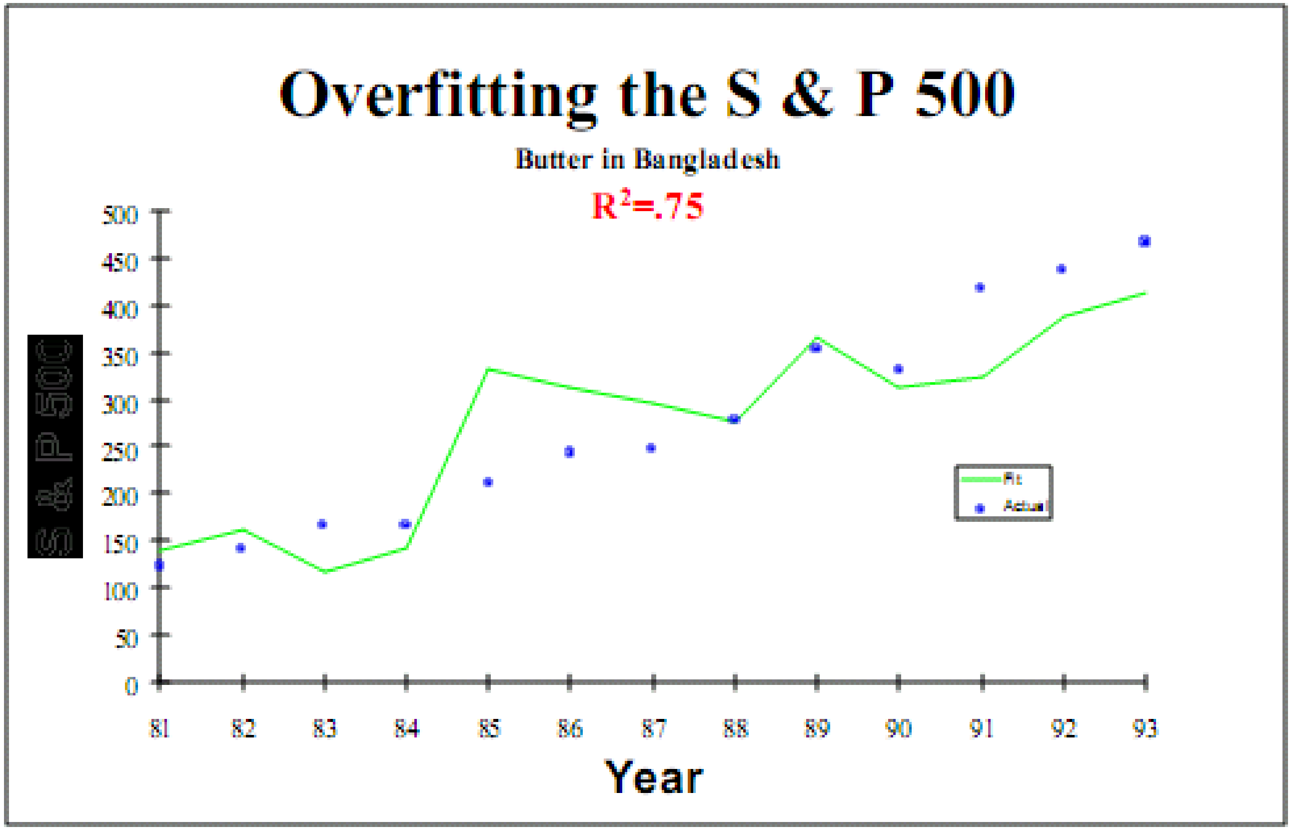
Correlation between two observables or series



Correlation and data-mining

We'll use the annual closing price of the S&P 500 index for the ten years from 1983 to 1993, shown in the chart below.



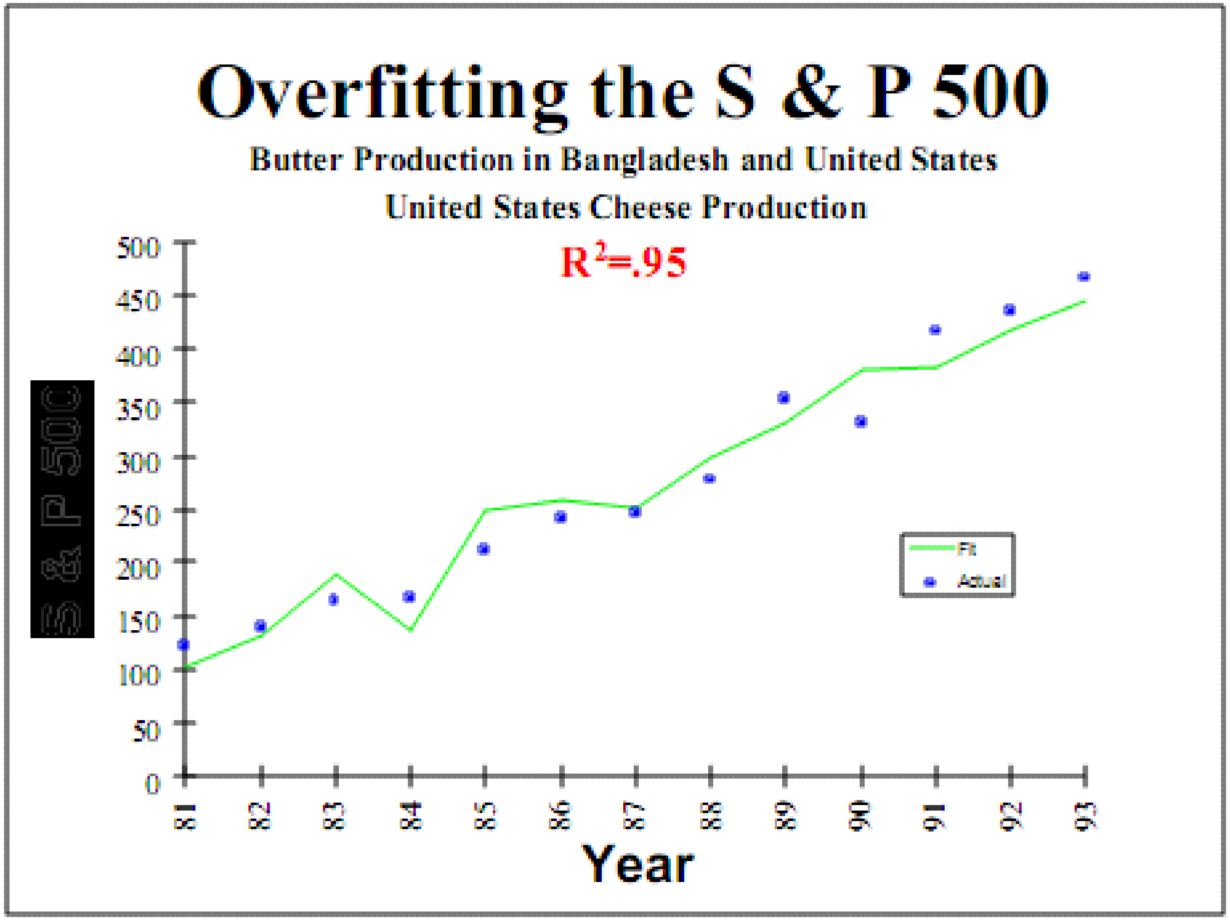


Stupid Data Miner Tricks: Overfitting the S&P 500

David J. Leinweber

THE JOURNAL OF INVESTING

Spring 2007

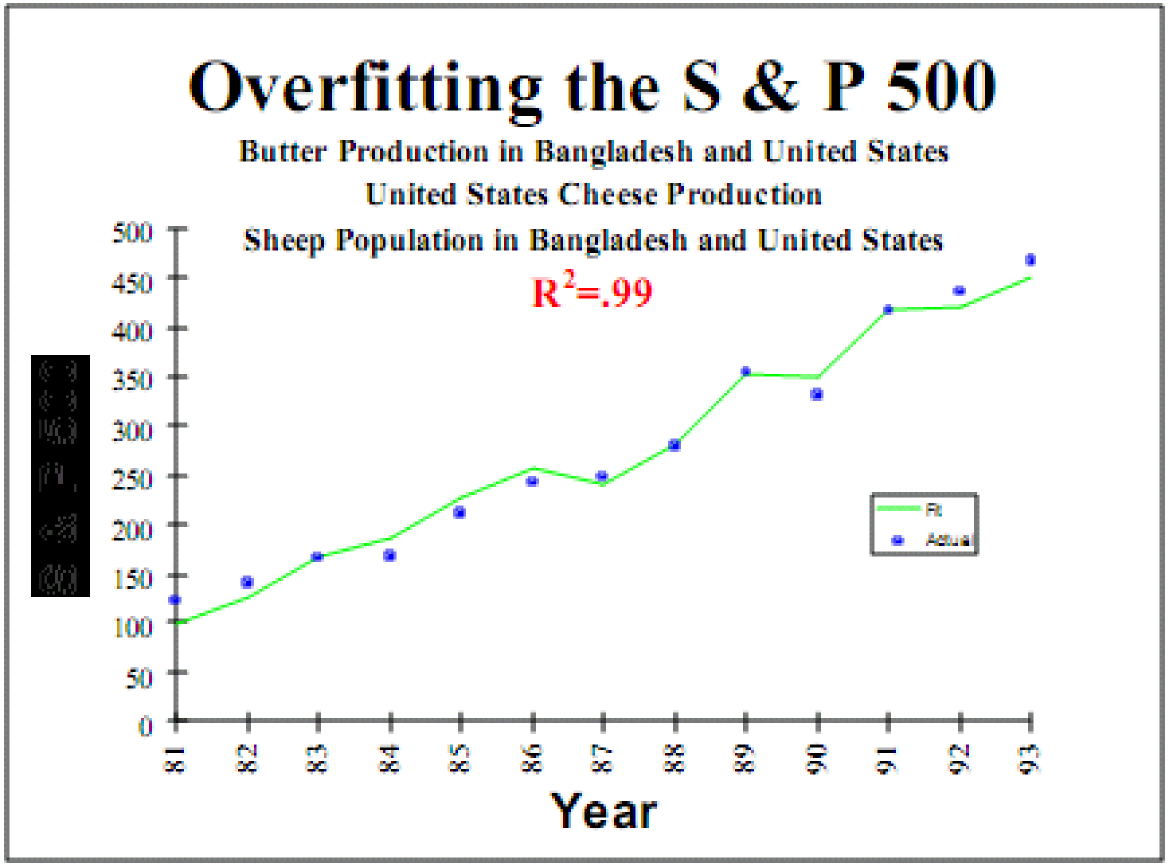


Stupid Data Miner Tricks: Overfitting the S&P 500

David J. Leinweber

THE JOURNAL OF INVESTING

Spring 2007



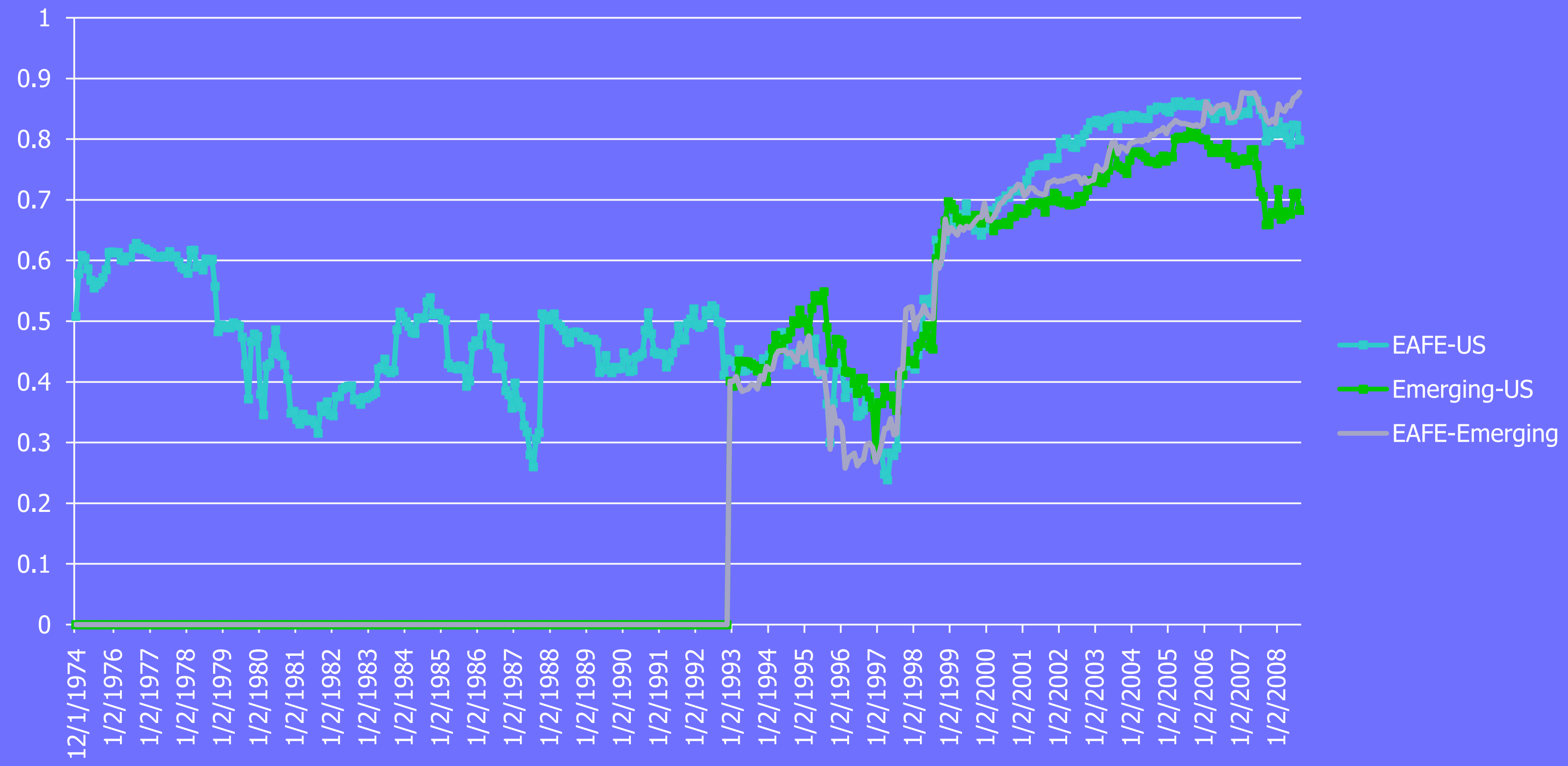
Stupid Data Miner Tricks: Overfitting the S&P 500

David J. Leinweber

THE JOURNAL OF INVESTING

Spring 2007

Historical correlation between stockmarkets



Correlation coefficients between rolling 5-year series of monthly returns of the indexes MSCI-Barra EAFE (Europe, Australasia, Far East), MSCI-U.S. and MSCI-Emerging Markets.

Mixing

Order n correlation coefficient:

$$c_n(\varphi, \psi) := \int \varphi.\psi \circ T^n d\mu - \int \varphi d\mu \int \psi d\mu$$

Ergodicity implies

$$\frac{1}{N} \sum_0^{N-1} c_n(\varphi, \psi) \longrightarrow 0$$

Mixing requires that

$$c_N(\varphi, \psi) \longrightarrow 0$$

namely φ and φ ∘ T^n become independent of each other as $n \to \infty$

Strong vs. weak mixing

- *Strongly mixing* systems are such that for every E, F, we have $\mu(T^n(E) \cap F) \rightarrow \mu(E)\,\mu(F)$ as n tends to infinity; the Bernoulli shift is a good example. Informally, this is saying that shifted sets become asymptotically independent of unshifted sets.

- *Weakly mixing* systems are such that for every E, F, we have $\mu(T^n(E) \cap F) \rightarrow \mu(E)\,\mu(F)$ as n tends to infinity *after excluding a set of exceptional values of n of asymptotic density zero*.

- Ergodicity does not imply $\mu(T^n(E) \cap F) \rightarrow \mu(E)\,\mu(F)$ but says that this is true for Cesaro averages:

Mixing of hyperbolic automorphisms of the 2-torus (Arnold's cat)