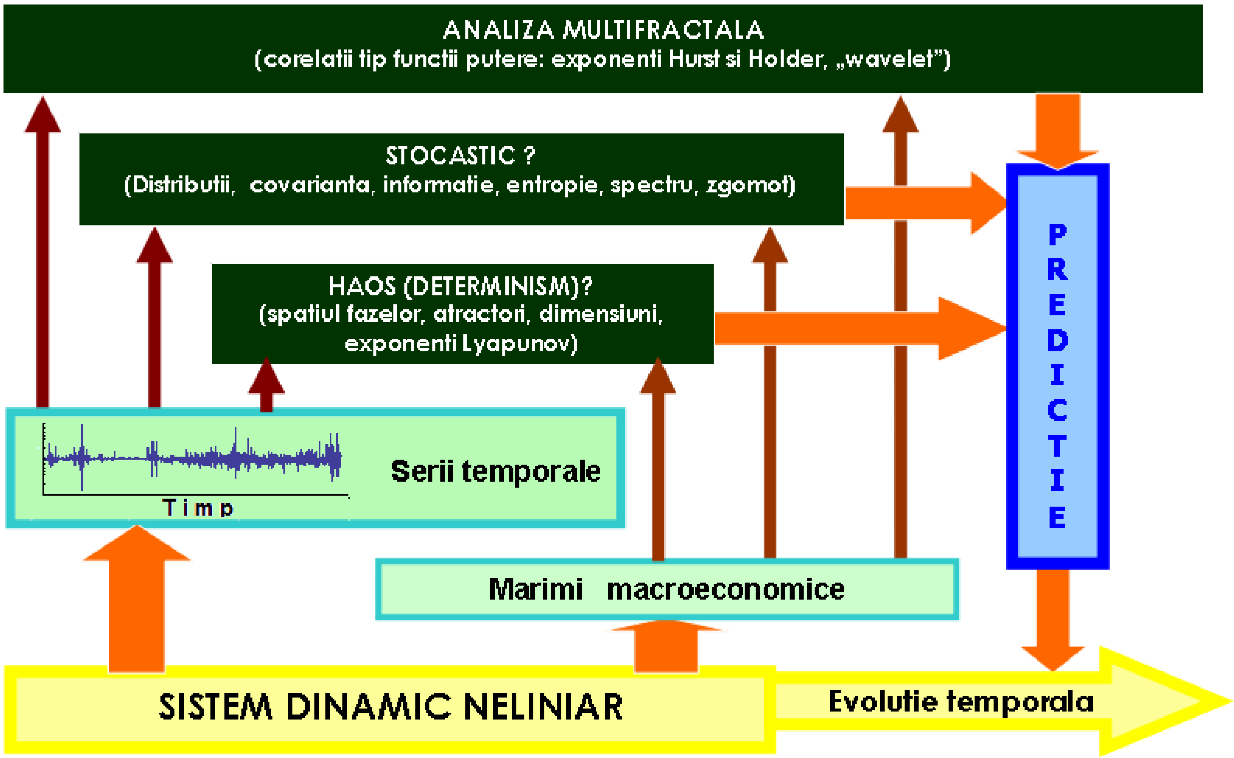
Elemente de econo-fizică

Subiectul este la intersecţia a trei domenii, anume Fizica proceselor neliniare, Statistică matematică, şi Teorie economică, aplicate în sectorul financiar-monetar. În ţările occidentale, cu economii avansate, această interdisciplinaritate a căpătat deja nume propriu, tradus şi în limba romană: econo-fizică.

În particular, analiza seriilor temporale ale indicatorilor economici, cursurilor de schimb sau indicilor bursieri este larg răspândită și utilizată pentru predicția și planificarea afacerilor.

Cuprins

I. Metodologia de studiu a seriilor temporale



Graficul seriei temporale

Graficele unor serii temporale sunt indicate in figurile de mai jos, anume cursul de schimb leu (ROL, valori denominate) – dolar american (USD), respectiv indicele bursier agregat BET al bursei de valori Bucuresti (BVB), ambele in functie de timp.

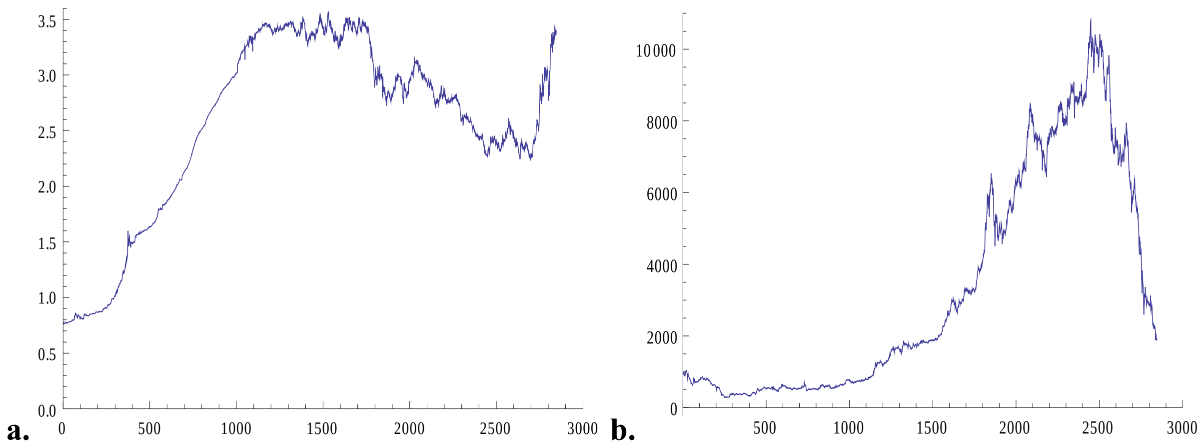


Fig.1 Cursul de schimb ROL-USD (a) şi indicele bursier agregat BET al bursei de valori Bucuresti (b), valori nominale (abscisele în zile, intre 19 sept.1997 - 28 feb. 2009, zile lucrătoare, 2846 valori).

În funcție de scopul fixat, inspecția vizuală a graficelor poate revela cel mai usor intervalele unde s-ar putea concentra studiul. Spre exemplu, intervalul 0-1100 zile din cursul de schimb ROL-USD este marcat de o creștere puternică, ce poate fi pusă în legătura cu creșterea indicelui prețurilor de consum din aceeași perioadă. Mai mult, în acelasi interval, indicele bursier manifestă o acalmie asemănătoare cu cea a cursului de schimb, dacă ne referim la amplitudinile micilor variații oscilatorii. Ascensiunea indicelui BVB începe înaintea deprecierii cusului de schimb, ceea ce ar putea justifica căutarea unei relații cauzale între cele două mărimi, pe lânga analiza de corelație.

Inspecția vizuală are însa limite, deoarece nu este sensibilă decât la relații liniare și figuri asemenea (în sens geometric), fiind mai putin sensibilă, sau deloc, la dependențe neliniare. Spre exemplu, în figurile de mai jos. cursul ROL-USD este dat în valori nominale și în valori logaritmate. După cum se observă, aplicarea unei transformari logaritmice conduce la punerea în evidență a unor caracteristici din zone diferite ale celor două grafice. Cele două forme de undă nu mai sunt asemenea, în sensul geometric. De aceea, de multe ori sunt necesare tehnici mai sofisticate de analiză pentru analiza curbelor.

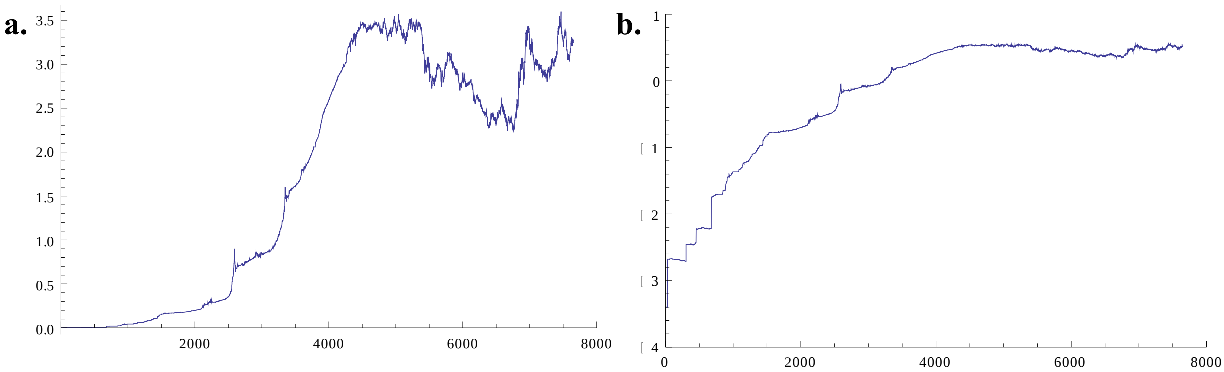


Fig.2 Cursul de schimb ROL-USD intre 1 ian.1990 - 31 dec. 2010, (a) valori nominale, si (b) valori logaritmate.

Semnificația valorilor logaritmice

Variații relative – randament (dobânda)

Aăa cum s-a arătat anterior[1], pentru a putea fi studiate, fiecărei proprietăți a unui sistem fizic i se asociază câte o mărime fizică măsurabilă. Pentru a putea compara sistemele între ele, unitățile de măsura sunt standardizate. În cazul sistemelor și proceselor economice, o astfel de standardizare lipsește. Prin urmare, compararea indicilor bursieri la Bucuresti și la Frankfurt nu este simplă, după cum nu este simplu sa se nici compare prețul unui dolar US, exprimat în lei, respectiv în Euro.

Așadar, cum trebuie prelucrate măsurătorile brute, pentru a extrage informația relevantă din punct de vedere al teoriilor economice?

[1] A se vedea fascicula *Introducere*.

O variație a prețului

$$\Delta P(n_1) = P(n_1+1) - P(n_1) = 0{,}002 \text{ ROL}$$

în jurul abscisei n_1= 452 (marcată cu săgeata în figura de mai jos) înseamna o variație relativă a prețului

$$\left.\frac{\Delta P}{P}\right|_{n_1} = \frac{0{,}002}{0{,}004} = 50\%.$$

Aceeași variație absolută, în jurul abscisei n_2= 4426, înseamna o variație relativă

$$\left.\frac{\Delta P}{P}\right|_{n_2} = \frac{0{,}002}{3{,}498} = 0{,}057\%,$$

adică de aproape o sută de mii de ori mai mare decât în cazul precedent!

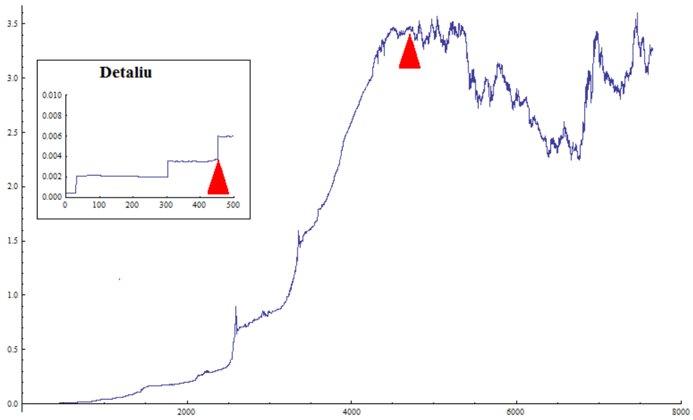


Fig.3 ROL-USD

Din punct de vedere economic, *variația relativă* este cea care are semnificație, din motivele enumerate mai jos.

i/ "Ceteris paribus[2]", o anumită variație relativă a valorii mărfii, aici unitatea monetară USD, înseamnă variații relative egale, indiferent de moneda de exprimare.

Exemplu

Fie cursul 2,80ROL=1\$=0,68€, și variația relativă a valorii unui USD de +1%, adică o crețtere de la 1\$ la 1,01\$. Noua valoare, exprimată în ROL, este 2,828ROL, adică o variație $\frac{2{,}828-2{,}800}{2{,}800} = 1\%$. Exprimată în Euro, noua valoare este 0,6868€, variația fiind $\frac{0{,}6868-0{,}68}{0{,}68} = 1\%$. Prin urmare, dacă toate celelalte condiții se mențin neschimbate, variațiile relative sunt identice.

Particularitatea sistemelor economico-financiare constă în aceea că, exprimate în monedă proprie, variațiile relative sunt diferite, consecința a unor conditii endogene, specifice.

ii/ Variația relativă este legată de randamentul investiției, motiv pentru care variația relativă se mai numeste randament[3] sau "Return" (*Ret*). Acest randament măsoară eficiența unei investiții. În cazul împrumuturilor financiare, randamentul este echivalent cu dobânda. Randamentul rezultat din modificarea prețului de la $P(t)$ la momentul t la $P(t+\mathrm{d}t)$ la momentul t+dt este

$$Ret(t) = \frac{P(t+\mathrm{d}t) - P(t)}{P(t)}.$$

[2] Toate celelalte condiții se mențin neschimbate.
[3] De la englezescul "return".

În variabile numerice,

$$Ret(n) = \frac{P(n+1)}{P(n)} - 1.$$

Dacă valoarea este astăzi de 100 (unități convenționale), în cazul unei dobânzi de 0,01% pe zi, valoarea devine 100,01 după prima zi, deoarece (100,01−100)/100=0,01%. Așadar, randamentul zilnic este de 0,01%, identic cu dobânda.

iii/ Variațiile relative elimină tendințele introduse, spre exemplu, de inflație. Dacă, pe un interval, este îndeplinită condiția

$$Ret(t)=constant,$$

consecința este că, rezolvând ecuația diferențială, rezultă o variație exponențială a prețului:

$$P(t)=P(0)\cdot e^{at},$$

unde $constant=a$. Dacă $a>0$, atunci tendința este de creștere exponențială a prețului.

Exemplu

În figura alăturată este dat cazul ROL-USD în perioada de inflație puternică 1992-2000, unde este evidentă comportarea de tip exponențial, cu constante diferite, pe porțiunile delimitate de liniile verticale. Cu toate acestea, indiferent de prețul nominal, variația relativă a acestuia este constanta (pe porțiuni). Punctele de racordare indică, probabil, intervenții ale Bancii Centrale, care, din timp in timp, au temperat deprecierea exponențială a cursului de schimb.

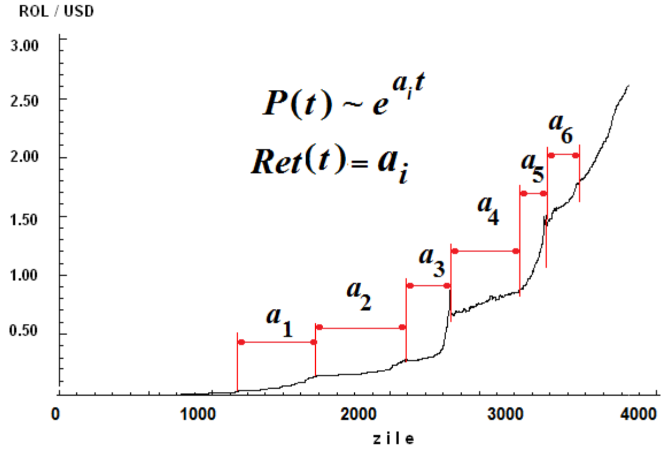


Fig.4 ROL-USD în perioada inflaționistă: indiferent de valoarea nominală a pretului, variația relativă a acestuia ramâne aproximativ constantă pe porțiuni

Valori logaritmate

Matematic, o variantă elegantă de a obține ușor variatiile relative este de a lucra cu *logaritmul* valorile nominale:

$$p(t)=\log P(t).$$

Se observă ușor că

$$Ret(t) = \frac{dp(t)}{dt} = \frac{d\left(\log P(t)\right)}{dt}$$

este echivalent cu

$$Ret(t) = \frac{1}{P(t)}\frac{dP}{dt}.$$

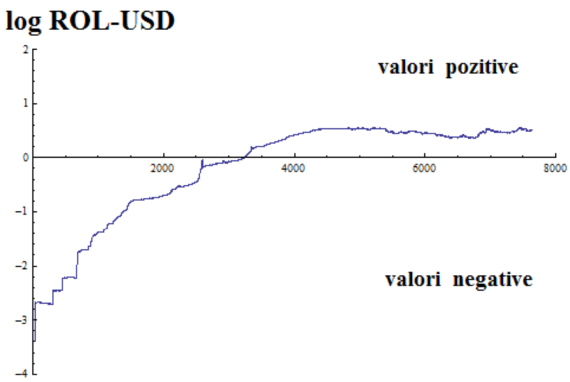


Fig.5 Logaritmarea extinde posibilitatea lucrului cu valori negative ale prețurilor

Un alt avantaj al lucrului cu valori logaritmate este acela că extinde posibilitatea lucrului în domeniul prețurilor negative. Formal, în anumite tehnici de simulare, unde, de exemplu, se ține cont de efectul inovarii, sau al denominărilor, există posibilitatea obținerii unor prețuri negative. Acest lucru nu are sens pentru valorile nominale, dar are sens pentru valorile logaritmate (Fig.5).

În sfârșit, după cum s-a arătat anterior (Fig.1a) transformarea neliniară logaritmică servește la punerea în evidență a unor caracteristici din zone depărtate în timp, asigurănd *comparabilitatea* variațiilor la scale mari de timp (zeci de ani), peste distorsiunile introduse de inflație, deflație, variații sezoniere, denomināri etc.

II. Caracterizarea seriilor temporale

În lucrarea de față, ne limităm la a caracteriza seriile din patru puncte de vedere: i/ spectrul de frecvențe al seriei, ii/ distribuția valorilor seriei, iii/ corelațiile dintre valorile seriei, iv/ fractalitatea.

Spectrul de frecvențe (Fourier)

Eșantionarea (Teorema Shannon-Nyquist)

În imensa majoritate a cazurilor, semnalele se procesează digital, motivul principal fiind acela al disponibilitatii, aproape exclusive, a metodelor numerice de prelucrare. În afara filtrărilor digitale, cele mai complexe prelucrări sunt codarea și criptarea: cea dintâi se foloseste pentru a proteja transmisiile de date la perturbaţii, în cazul comunicațiilor comerciale, iar cea de-a doua este tehnică matematică, bazată pe analiză probabilistică, utilizată pentru securizarea informațiilor, în special în cazul informațiilor financiar bancare. Din acest motiv, semnalele trebuie *eşantionate*.

Efectul eșantionării este acela că semnalele analogice se transformă în semnale digitale, sub formă de serii temporale:

$$x(t) \to x(nT) \sim x(n).$$

Teorema Shannon-Nyquist[4] - indică modul cum trebuie eşantionat un semnal a cărui energie este cuprinsă, de exemplu, în proporţie de 95%) în banda $\omega \in (0, \omega_{max})$:

Intervalul de eşantionare Δt_{esant} – marcat cu sageți în figura alăturată – trebuie să fie *mai mic decât jumătate din perioada componentei cu frecvența cea mai mare* din spectrul semnalului:

$$\Delta t_{esant} \le \frac{\pi}{\omega_{max}} = \frac{1}{2 f_{max}}, \text{ unde } \omega_{max} = 2\pi f_{max}.$$

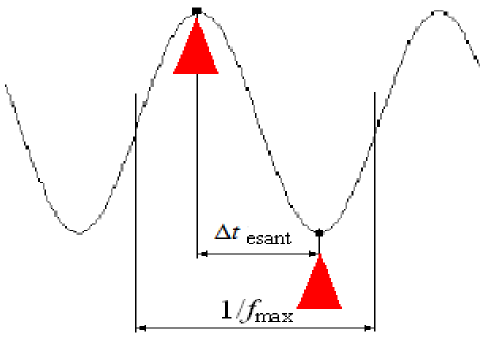


Fig. 6 Interpretarea fizica a condiției Nyquist

[4] A se vedea fascicula *Oscilații* și http://en.wikipedia.org/wiki/Nyquist%E2%80%93Shannon_sampling_theorem

Interpretarea fizică este aceea că, la limita impusă de condiţia Shannon-Nyquist, frecvenţa de eşantionare este suficient de mare – echivalent, eşantioanele sunt suficient de dese – încât să poată prinde variaţiile temporale cele mai rapide: in Fig.6, este vorba de valorile marcate, adică cele maxime şi minime ale semnalului sinusoidal cu pulsatia ω_{max}.

Evident, frecvenţele mai înalte decât $2\pi f_{max}=\omega_{max}$ din spectrul semnalului (reprezentând, în exemplul nostru 5% din energia sa) nu pot fi puse în evidenţă. De aceea, semnalul reconstituit nu este identic cu cel original.

Analiza în frecvență

Dintr-un semnal de durată T_{total}, care se eşantionează cu pasul temporal Δt_{esant}, se obțin N eşantioane în domeniul timp, $T_{total} = N \cdot \Delta t_{esant}$. Efectuând o transformare Fourier asupra seriei $x(n\cdot\Delta t_{esant})$ cu ajutorul programelor matematice, se obțin $N/2$ eşantioane ale seriei $X(k\cdot\Delta f)$, cu pasul de frecvență Δf, unde corespondențele dintre reprezentarile timp-frecvenţă sunt indicate mai jos.

i/ *Rezoluția spectrală*, sau pasul spectral Δf, ca intervalul dintre două eşantioane consecutive, exprimat în unităţi de interval spectral

$$\Delta f = \frac{1}{T_{total}},$$

ii/ *Frecvenţa maximă* detectabilă

$$f_{max} = \frac{1}{2\cdot\Delta t_{esant}},$$

căreia îi corespunde perioada minimă detectabilă $2\Delta t_{esant}$.

iii/ *Frecvența minimă* detectabilă, în afară de componenta continuă, de frecvență zero

$$f_{min} = \frac{1}{T_{total}}.$$

Corespunzător, perioada maximă detectabilă este T_{total}. Frecvența minimă detectabilă este egală cu pasul spectral.

iv/ Numărul de eșantioane Q din spectrul de frecvență este *jumătate* din numărul de eșantioane N din domeniul temporal (dar informația conținută este aceeași!):

$$Q = \frac{f_{max}}{\Delta f} = \frac{N}{2}, \qquad \text{(f)}$$

În consecință, unul dintre avantajele lucrului în domeniul frecvență este acela că procesarea este mai rapidă.

Transformarea Fourier

Semnalele continue $x(t)$ se pot analiza în domeniul frecvenţelor $\omega=2\pi f$, utilizând *transformarea Fourier*[5]

$$X(\omega)=\int_{-\infty}^{\infty} x(t)e^{-\mathbf{i}\omega t}\mathrm{d}t\,,\qquad X(k)=\sum_{n=0}^{N-1} x(n)\cdot e^{-\mathbf{i}2\pi k\frac{n}{N}}\,,\ k=0,\dots,N-1,$$

unde $X(\omega)$, $X(k)$ sunt transformatele Fourier ale semnalului, în varianta continuă, respectiv discretă, iar $\mathbf{i}=\sqrt{-1}$.

Relaţia reprezintă o sondare a semnalului $x(t)$ cu funcția armonică complexă $e^{-\mathbf{i}\omega t}=\cos(\omega t)-\mathbf{i}\sin(\omega t)$, denumită și *nucleul* transformării, și arată în ce măsură componenta sinusoidală, de frecvenţa ω, se regăseşte în semnalul x (atât ca modul, cât şi ca fază).

Modelul matematic consideră semnale nemarginite în timp şi frecvenţă, ceea ce nu avem, practic, niciodată. Toate semnalele reale sunt finite în timp. De exemplu, doar un semnal sinusoidal infinit în timp are frecvenţa perfect precizată. Dacă, însa, semnalul are durată finită, cu un început şi un sfarşit, frecvenţa lui nu mai este perfect precizată, el având mai multe frecvente, sau o bandă de frecvenţe. Încercând să precizăm (localizăm) o funcţie în timp, spectrul ei se va lărgi în frecvenţă, şi reciproc, încercarea de a preciza intervalul spectral necesită timpi de măsură mai lungi[6].

Ca interpretare fizică, $\left|x(n\Delta t_{\text{esant}})\right|^2\Delta t_{\text{esant}}$ este energia pe intervalul de timp Δt_{esant} din jurul momentului $n\Delta t_{\text{esant}}$; analog, $\left|X(k\Delta f)\right|^2\Delta f$ este energia pe intervalul spectral Δf din jurul componentei $k\Delta f$. Totusi, trebuie remarcat că $\left|x(n)\right|^2$ are dimensiune de putere (J/s), pe când $\left|X(k)\right|^2$ are dimensiune de acțiune (Js). Graficul, în coordonate $(k,\left|X(k)\right|^2)$, constituie *spectrul* lui $x(n)$.

Urmare a conservării energiei, rezultă[7]:

$$\int_0^{T_{\text{total}}}\left|x(t)\right|^2\mathrm{d}t=\int_0^{\omega_{\max}}\left|X(\omega)\right|^2\mathrm{d}\omega\,.$$

Varianta discretă

$$\sum_{n=0}^{N-1}\left|x(n)\right|^2=\sum_{k=0}^{N-1}\left|X(k)\right|^2\,,$$

se obține normând pașii de eșantionare la unitate; pentru eșantioanele temporale

$$\sum_{n=0}^{N-1}\left|x(n)\right|^2=\sum_{n=0}^{N-1}\left|x(n\Delta t_{\text{esant}})\right|^2\cdot\Delta t_{\text{esant}}\Big|_{\Delta t_{\text{esant}}=1}\,.$$

[5] Pentru detalii, a se vedea
[6] În esenţă, compromisul timp-frecvenţă ilustrează relaţiile de incertitudine (Heisenberg). Există şi tehnica de localizare simultană în timp şi în frecvenţă, cunoscută sub denumirea de transformarea "wavelet".
[7] Demonstrarea relației este cunoscută drept *teorema lui Parseval*.

Dacă seria x este o serie financiară, atunci modulul pătrat al valorii eșantionului nu are un corespondent interpretabil, ca în cazul semnalelor fizice. Spre exemplu, dacă $x(n)$ este prețul (mediu) al unei unități de marfă în ziua n, exprimat în unități monetare, atunci pătratul acesteia nu are sens economic.

Transformarea inversă permite reconstituirea semnalului iniţial

$$x(t)=\frac{1}{2\pi}\int_{-\infty}^{\infty}X(\omega)e^{\mathrm{i}\omega t}\mathrm{d}\omega,\ x(n)=\frac{1}{2\pi}\sum_{k=0}^{N-1}X(k)\cdot e^{\mathbf{i}2\pi \boldsymbol{k}\frac{\boldsymbol{n}}{\boldsymbol{N}}},\ n=0,\dots,N-1$$

unde factorul $1/2\pi$ este un factor de scală, care asigură regăsirea aplitudinii originale. Avantajul lucrului în domeniul frecvență este acela că, spre exemplu, ecuațiile diferențiale liniare devin ecuatii algebrice, a căror soluție este mai simplu de găsit.

Din punct de vedere matematic, atât transformarea directă, cât și cea inversă, transformă, în general, mărimi complexe în marimi, de asemenea, complexe. Din cauză ca $x(n)$ sunt însă valori *reale*, eșantioanele corespunzatoare $X(k)$ sunt *conjugate* două câte două:

$$X(k)=X^{*}(N-k),$$

asadar

$$|X(k)|=|X^{*}(N-k)|. \tag{f-1}$$

Acesta este motivul pentru care *jumătate* din eșantioanele din domeniul frecvență sunt suficiente pentru reconstrucția semnalului temporal:

$$Q=\frac{f_{\max}}{\Delta f}=\frac{N}{2}, \tag{f}$$

unde Q este numărul de eșantioane din spectrul de frecvențe.

Observație

Panta spectrului, aproximat, în coordonate dublu logaritmice ($\log k$, $\log|X(k)|^2$), cu o dreaptă,

$$\log|X(k)|^2 \sim panta\cdot\log k,$$

oferă informații importante pentru analizele incrucișate.

Exemplu

Fie cursul de schimb ROL-USD 1 ianuarie 2001-31 martie 2011, N=3740 zile (aproximativ 10 ani). Deoarece diferența dintre valorile $|X(k)|^2$ este foarte mare, s-a optat pentru reprezentarea logaritmului acestora, anume $\log|X(k)|^2$.

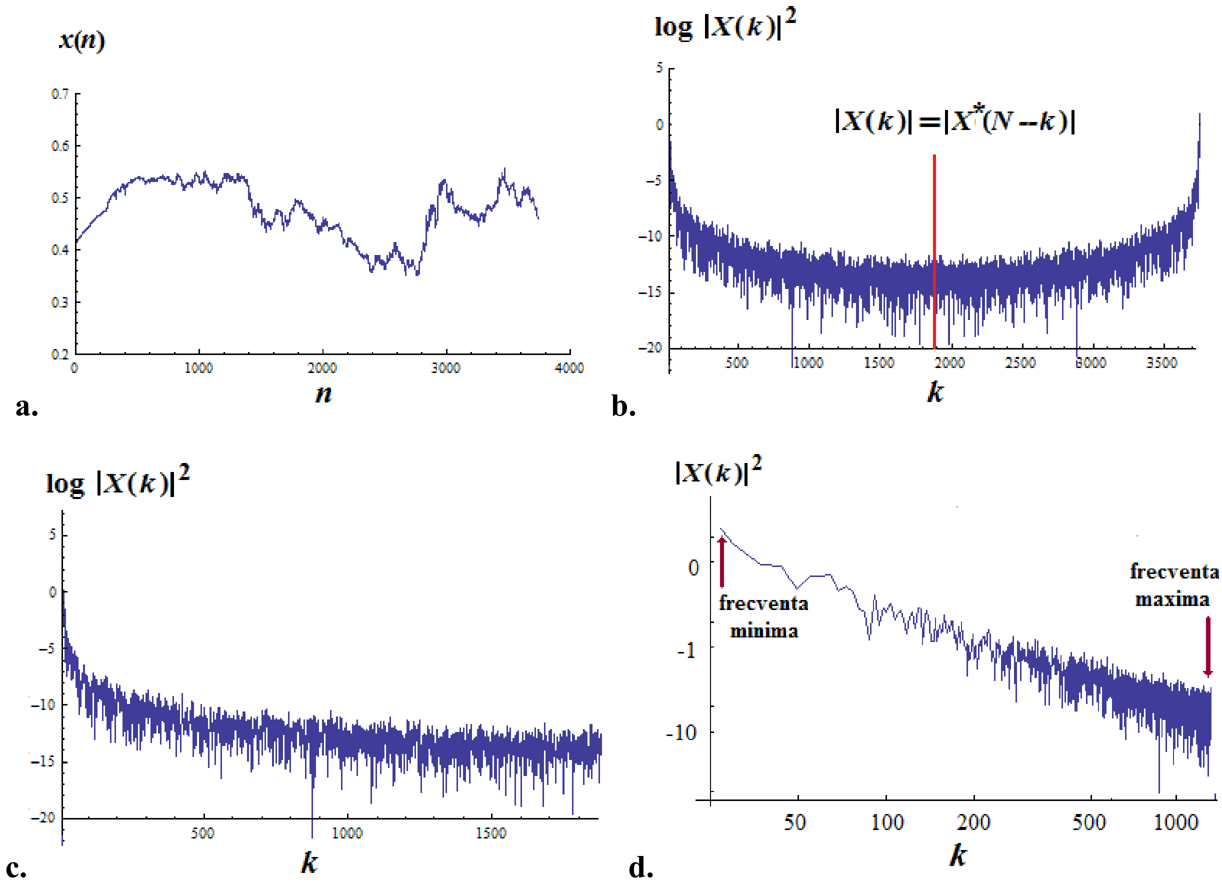


Fig.7. Cursul de schimb ROL-USD (valori logaritmice), 1 ian.1996- 31 mar.2011 (a), și spectrul sau Fourier, la scară semi-logaritmică: se observă simetria valorilor față de mijlocul $N/2$ (b), spectrul Fourier cu Q=1870 componente spectrale (c), spectrul Fourier la scara dublu logaritmică (d).

Corespondențele dintre domeniile timp și frecvență sunt ilustrate în tabelul următor.

Tabelul II

Timp	**Frecvența (ciclu)**
Interval de eșantionare Δt_{esant}=1 zi	Frecventa maximă observabilă: f_{max}= 0,5 zi^{-1} (2 zile)
Numar eșantioane N=3740	Frecvența minimă observabilă$^{*)}$: f_{min}= 1/3740 zi^{-1} (3740 zile)
Durata seriei T_{total}=N zile	Rezoluția în frecvență: Δf= 1/3740 zi^{-1}
	Număr eșantioane: Q=1870

*) Exclusiv componenta continua

Filtrarea

Prin filtrare se elimină anumite componente din seria originală, precum tendințele (crescătoare, sau descrescătoare), ciclicitățile (cicluri economice, de afaceri, sezoniere etc.), sau anumite tipuri de zgomote. În felul acesta, se poate analiza ponderea acestor componente în structura prețului, sau a mărimii care constituie seria temporală.

Filtrarea se poate face fie în timp, fie în frecvență. Dintre filtrarile în timp, cele mai cunoscute și mai simple sunt eliminarea tendințelor, derivarea termen cu termen, auto-regresiile, și medierea cu fereastra mobilă[8]. Ultimele două, combinate, sunt utilizate pe scara largă în modelele ARMA[9].

Exemplu: staționarizarea seriilor

Multe metode de analiză impun lucrul cu serii staționare[10]. Una dintre posibilitatile de a obține serii staționare, dar nu singura, este aceea de a elimina tendințele, fie prin filtrare liniară, fie prin tehnici neliniare. Tendințele din seriile financiare pot avea cauze diverse. Spre exemplu, politica monetară a bancii centrale poate conduce la existența unor tendințe usor crescătoare pe termen lung.

Reziduurile valorilor logarimice ROL-USD şi BET care staţionarizează seriile pe termen lung (>100 zile); tendinţele sunt extrase prin aproximare, prin metoda celor mai mici pătrate, cu polinoame de grad G=85; ceea ce ramane este seria de *reziduuri* notata ***Res***.

$$Res\ x(t) = \log x(t) - \sum_{0}^{85} a_k t^k \ , \qquad Res\ y(t) = \log y(t) - \sum_{0}^{85} b_k t^k \ .$$

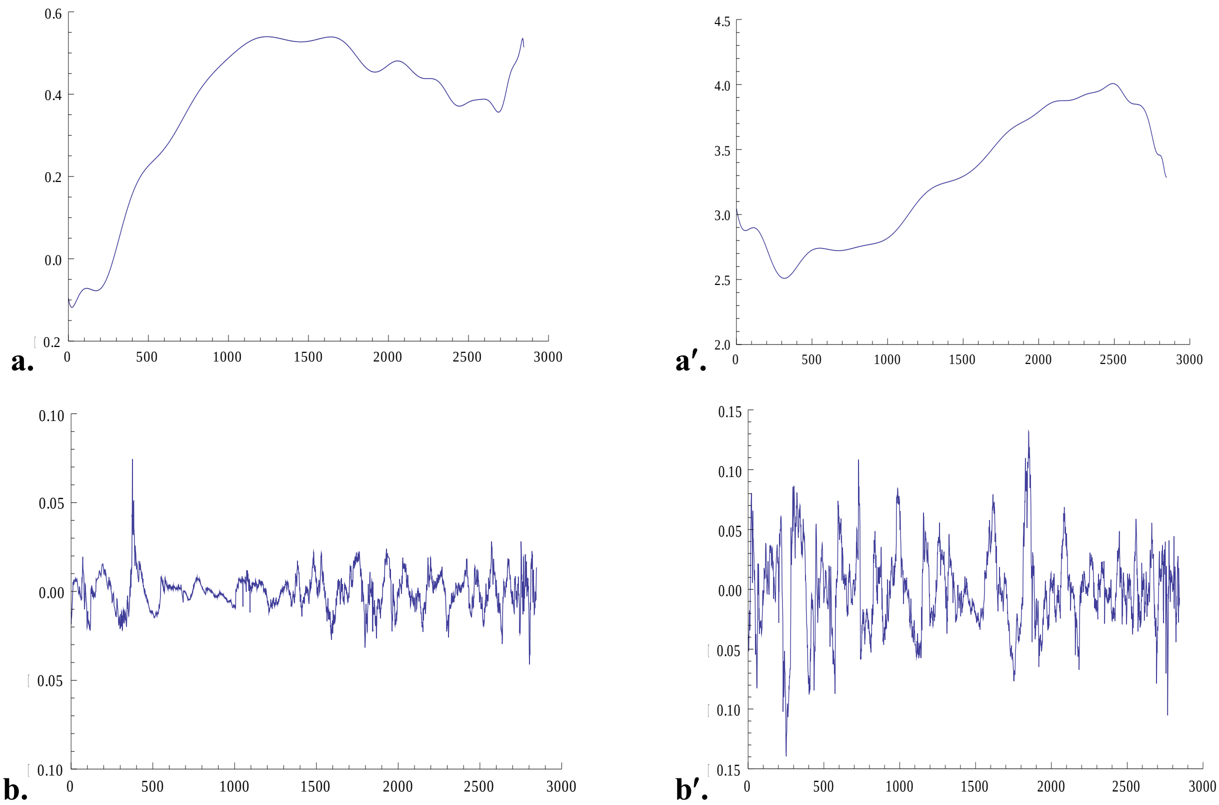


Fig.8 Tendinţa (a) şi reziduul (b) cursului de schimb ROL-USD, respectiv tendinţa (a′) şi reziduul (b′) indicelui bursier agregat BET (abscisa în zile).

[8] In limba engleza „moving average”.
[9] ARMA este acronimul de la „Auto-regressive Moving Average”.
[10] Vezi Cap.…

Pentru domeniul frecvență, filtrarea numerică, prin mascare, este cea care se utilizează cel mai frecvent. Aceasta constă în operarea modificărilor în spațiul Fourier cu condiția ca, dacă se operează asupra unui eșantion $X(k)$, să se facă aceeași operație și asupra conjugatului acestuia, conform relației (f).

Distribuții

O mulțime de N măsuratori succesive se poate pune sub forma unei serii temporale y_1, y_2,... y_N. Pentru orice astfel de serie se poate construi o histogramă, în felul următor: se partiționează intervalul în care apar valorile măsurătorilor într-un număr de I_N subintervale, și se numără valorile care pică în fiecare interval: fie acesta N_k, adică numărul de măsuratori ale căror valori sunt în intervalul cu numărul de ordine k. Frecvențele de apariție sunt N_k/N. Reprezentarea în coordonatele “subintervale”-„frecvențe de apariție” constituie histograma.

Exemplu

În exemplul de mai jos N=500, $y \in [-0,006; 0,006]$, n=60 cu rezoluție de 0,0002. Intervalul #1 este [–0,0060; –0,0058) și conține N_1=1 măsurători, intervalul #2 este [–0,0058, –0,0056) și conține N_2=0 măsurători..., intervalul #60 este [0,0058; 0,0060) și conține N_{60}=0 măsurători. Cele mai multe măsurători, anume 74, sunt în intervalul #32 [0,0002; 0,0004).

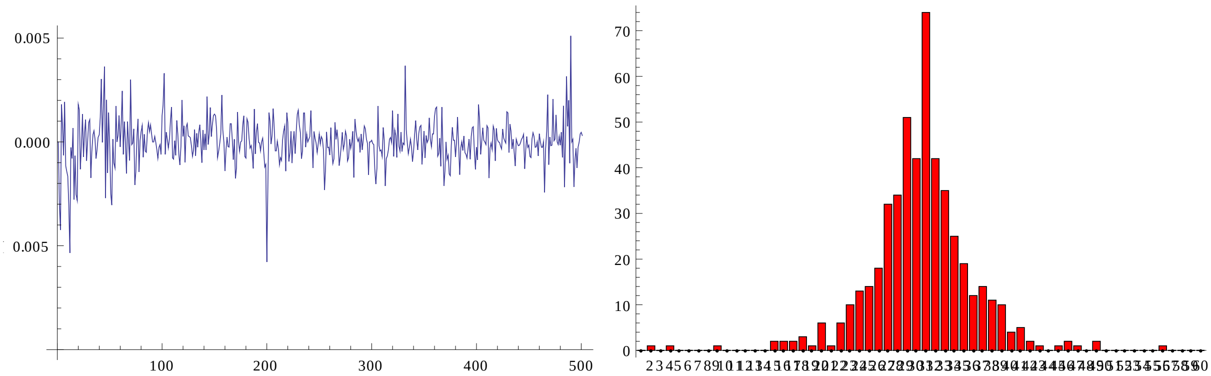


Fig.9 Seria temporală și histograma asociata

Densitatea de probabilitate („probability density function” - PDF)

Dacă $N \to \infty$, atunci histograma devine *distribuție*, în sensul că, pe abscisă, apare valoarea măsurată, iar pe ordonată, densitatea de probabilitate (PDF, notată Π), sub forma unei funcții continue. În acesta forma, densitatea de probabilitate nu are semnificație fizică. Ea capătă semnificație de probabilitate doar dacă se precizează intervalul în care se încadrează valoarea măsurată. Pentru exemplul anterior, probabilitatea ca mărimea măsurată să fie în intervalul [–0,0060; –0,0056) este

$$\mathcal{P}[-0,0060 \le y < -0,0056] = 0,0002 \cdot \sum_{k=1}^{k=2} N_k = (1+0)/500 = 0,002.$$

În aproximația continuă, probabilitatea ca variabila să fie cuprinsă în intervalul specificat se calculează cu suma continuă, simbolizată prin operatorul de integrare:

$$\mathcal{P}[-0{,}00060 \le y < -0{,}00056] = \int_{-0{,}00060}^{-0{,}00056} \Pi(y)\mathrm{d}y\,.$$

Momente

Momentul de ordin 1 al valorilor unei serii y de lungime N este *valoarea medie*:

$$m(1) = \frac{1}{N}\sum_{i=1}^{N} y_i\,.$$

Momentul de ordin "q" al unei distribuții provenită de la valorile unei serii y de lungime N se calculează cu relația:

$$m(q) = \frac{1}{N}\sum_{i=1}^{N}\left(y_i - m(1)\right)^q$$

În particular, momentul de ordin doi se numește *varianță*:

$$m(2) = \frac{1}{N}\sum_{i=1}^{N}\left(y_i - m(1)\right)^2,$$

iar rădacina pătrată a varianței este *abaterea standard* sau *fluctuația*:

$$\sigma = \left[\frac{1}{N}\sum_{i=1}^{N}\left(y_i - m(1)\right)^2\right]^{\frac{1}{2}}.$$

În cazul seriilor financiare, fluctuațiile se numesc *volatilități* și au semnificația *riscului mediu* al investiției.

Factori de formă: concentrarea "kurtosis" și asimetria "skewness"

Concentrarea (K) indică aglomerarea valorilor din vârfurile și cozile distribuției în raport cu centrul:

$$K = \frac{(N-1)(N+1)}{(N-2)(N-3)}\cdot\frac{m(4)}{\left(m(2)\right)^2}\,.$$

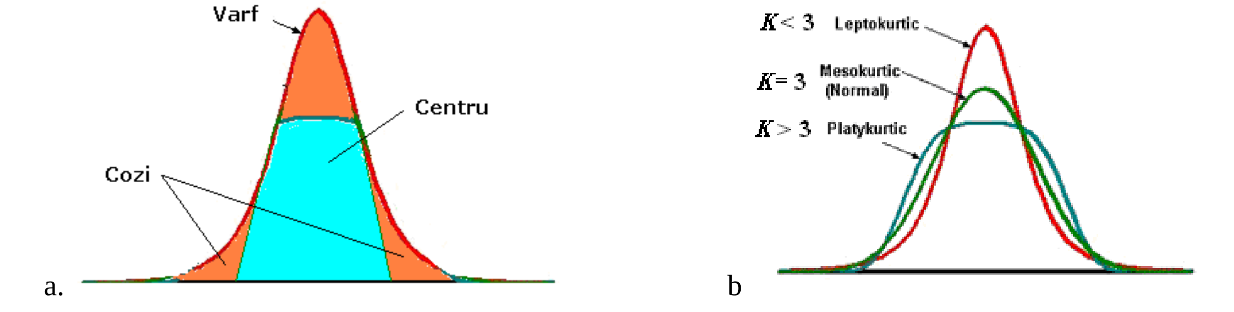


Fig.10 Concentrarea distribuției

O concentrare mai mare decât 3 indică o distribuție cu varf și cozi mai proeminente față de distribuția normală cu aceeași varianță. Dimpotrivă, o concentrare mai mică decât 3 indică o distribuție plată, cu cozi abrupte. Distribuția normală (Gaussiană) are K=3.

Asimetria (S) indică asimetria distribuției și se calculează cu relația:

$$S = \frac{\sqrt{N(N+1)}}{N-2} \cdot \frac{m(3)}{(m(2))^{3/2}} \; .$$

O asimetrie „pozitivă” S>0 indică o coadă mare spre dreapta; corespunzator, asimetria „negativă” S<0 indică o coadă mare spre stânga. Distribuția normală (Gaussiană) are S=0.

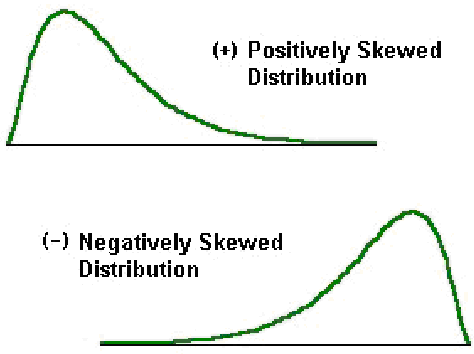


Fig.11 Asimetria distribuției

Definițiile anterioare nu au sens decât pentru forme de tip „clopot” ale distribuțiilor. O distribuție de tipul celei din Fig.12 este consecința unei nestaționarități a seriei temporale, în sensul ca media și varianța depind de timp.

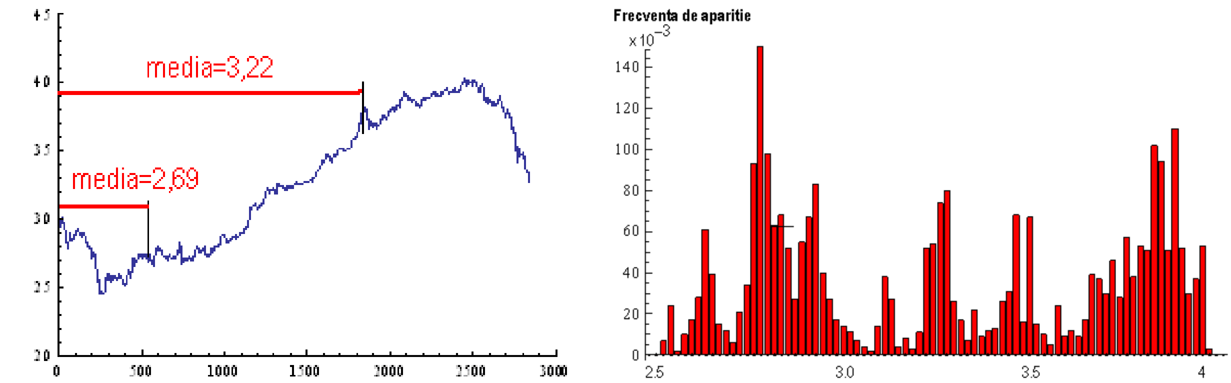


Fig.12

Pentru a putea compara distribuțiile, este necesar ca acestea să fie staționare, adică $m(1)$ si $m(2)$ să nu depinda de timp; din acest motiv, seriile se filtrează, pentru a le staționariza.

Exemplu: staționarizare prin transformare în serie de variații relative (randamente)

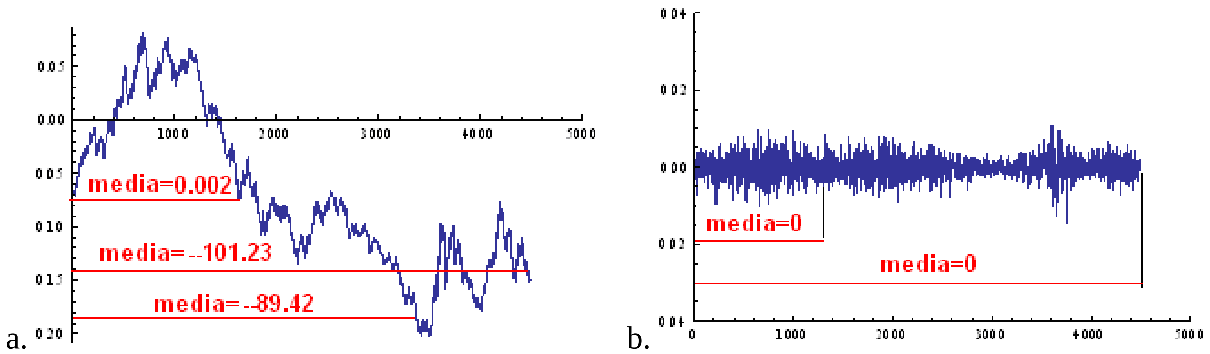


Fig.13 $p(t)$: serie nestaționară (a), $Ret(t)$: seria staționară a randamentelor (b) .

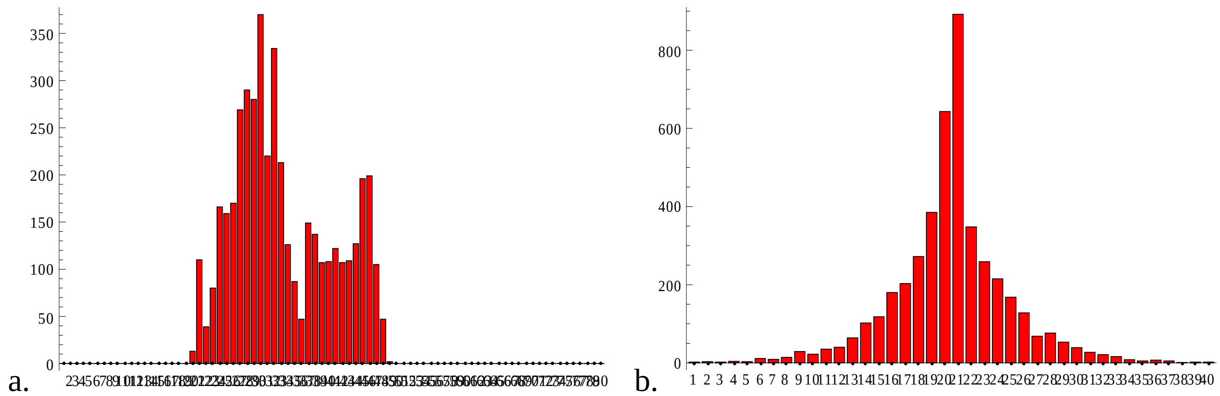


Fig.14 Distribuția valorilor $p(t)$ (serie nestaționară) (a), distribuția valorilor $Ret(t)$

O clasă importantă de distribuții nestaționare este generată de așa numitele drumuri aleatorii.

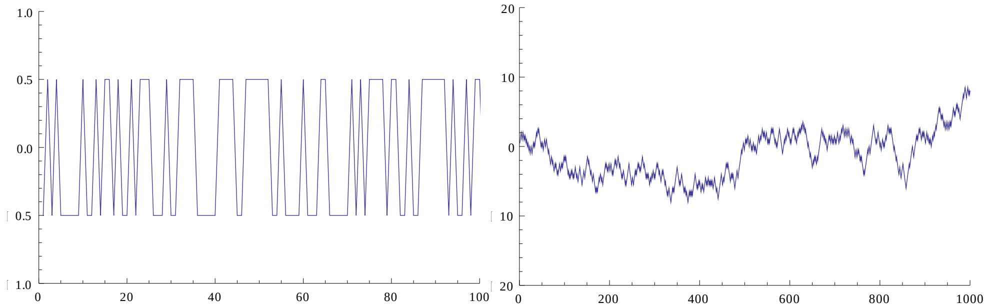
Drumuri aleatorii - "random walks"

Notiunea *drum aleatoriu* - "random walk" este extrem de des întâlnită în orice abordare a seriilor temporale, fiind importantă pentru înțelegerea legăturii cu distribuția Gaussiană. În figura de mai jos sunt ilustrate astfel de "drumuri", construite prin adunarea la valoarea curentă a numărului +1/2 sau −1/2, după cum ar apărea aceste numere într-un ipotetic experiment cu aruncarea unei monede cu două fețe, una inscripționată +1/2, cealaltă −1/2. Aruncarea repetată a monedei și înregistrarea valorilor succesive care se obțin constituie secvențe aleatoare compuse din valorile $\left\{-\frac{1}{2};+\frac{1}{2}\right\}$, care pot apărea cu probabilități egale, anume 0,5.

Analogia cu un drum aleator pornește de la ideea că, la fiecare aruncare a monedei, o persoană sa parcurgă, în lungul axei x, un pas de lungime 1/2, în sensul axei, dacă la experiment a ieșit valoarea +1/2, sau în sens contrar axei, dacă la experiment a ieșit valoarea −1/2. Șirul sumelor parțiale generează drumul aleator în coordonate (t, x). Ultima valoare din șir este coordonata (distanța față de origine) la care se găsește persoana, la sfârșitul secvenței de pași. Dacă se construiește un număr mare de astfel de secvențe identice, și se strâng toate aceste ultime valori, se constată că aproximează o distribuție Gaussiană. Aproximarea este cu atât mai bună, cu cât numărul de secvențe este mai mare.

Echivalent, dacă se consideră M secvențe independente de secvențe de lungime N fiecare, și se constituie matricea $M\times N$ cu elementele $a_{ij} \in \left\{-\frac{1}{2};+\frac{1}{2}\right\}, \quad i=1,...,N, \quad j=1,...,M$, sumarea pe coloane, după indicele j, va genera N sume ale căror valori sunt distribuite Gauss când $N\to\infty$.

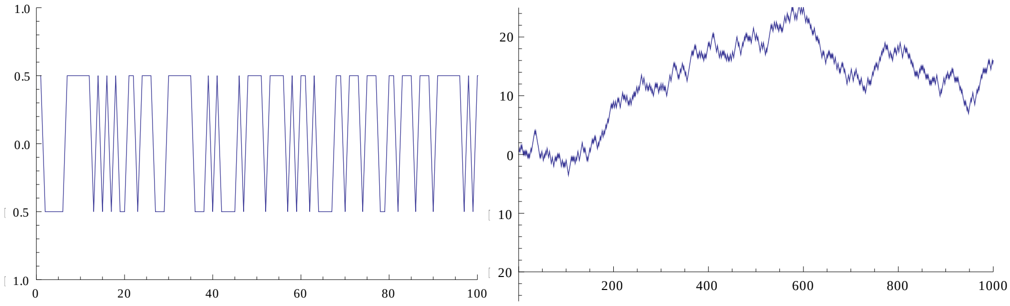
Acest lucru este indicat în figurile de mai jos.



Secvența 1 și drumul aleator generat, prin sumare, de sumele parțiale ale acesteia

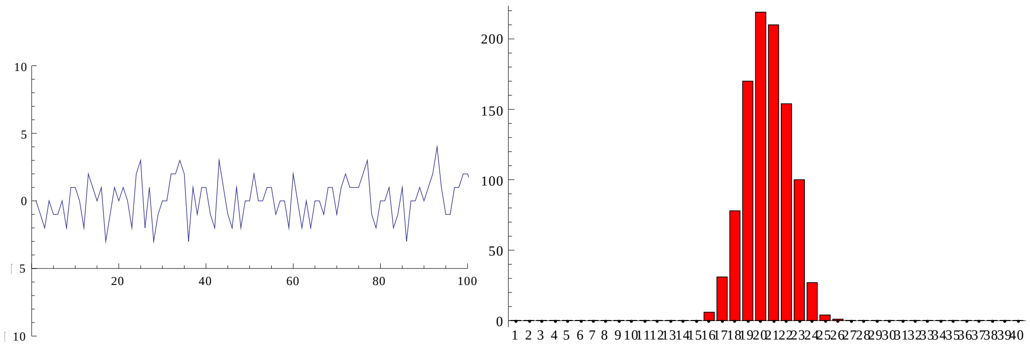
...

+



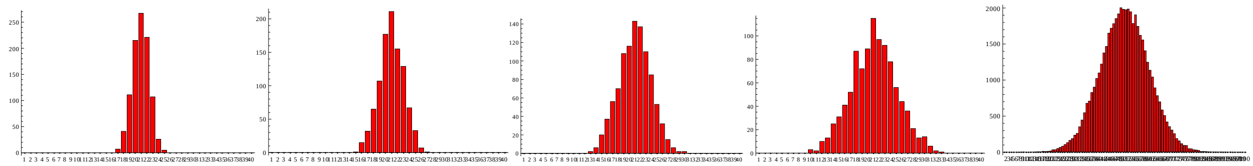
=

Secvența *M* și drumul aleator generat, prin sumare, de sumele parțiale ale acesteia



Suma dupa *M*=10000 a tuturor secvențelor și distribuția drumului aleator obținut, care tinde spre distibuția Gaussiană

Fig.15



Corelații

Corelațiile indică dacă valorile consecutive au legătura între ele[11]. Legatura se manifestă prin aceea că valoarea curentă poate depinde, într-un fel sau altul, de valorile anterioare. Aceeași distribuție poate fi formată din secvențe temporale distincte, consecință a regulilor după care se desfățoară procesele în interiorul sistemului care genereaza seria. Legătura poate să fie cauzală sau nu. Spre exemplu, să considerăm prețurile petrolului, benzinei și motorinei dintr-o anumită țară. Creșterea prețului petrolului *are drept cauză* creșterea costului obținerii benzinei, așadar între acestea există o legătură cauzală. Analog, există legătură cauzală și între prețul petrolului și prețul de cost al motorinei. Atât prețul benzinei, cât și al motorinei, cresc. Acestea sunt corelate, fară a exista o relație de tip cauză-efect între ele.

În cazul unor valori corelate, valoarea curentă depinde, după o funcție Φ, de valorile anterioare:

$$y(n) = \Phi\big(y(n-1),..., y(n-k)\big).$$

Dependența poate fi probabilistică, în care caz

$$\mathcal{P}\big[y(n)\big] = \mathcal{P}\big[y(n) \;|\; y(n-1),..., y(n-k)\big].$$

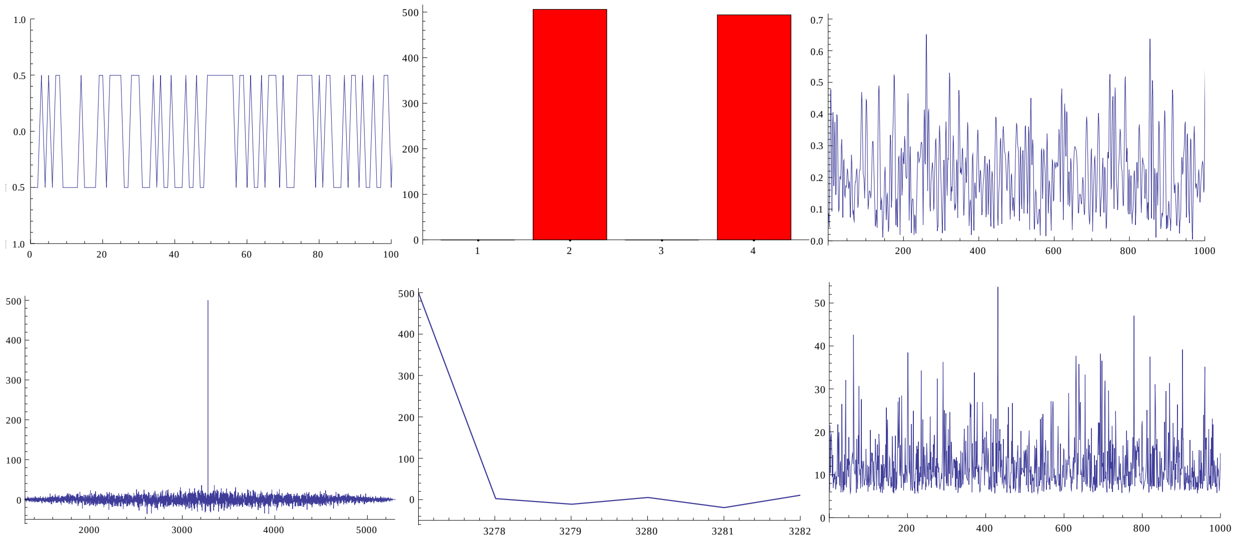
Din punct de vedere al seriilor financiare[12], funcţia de corelație la pasul τ nu este altceva decât o măsură statistică a tăriei cu care preţul actual seamănă cu preţul existent cu τ pași în trecut. Cu alte cuvinte, aceasta indică modul cum se poate prezice valoarea viitoare, pe baza unei singure măsurători. Suma de funcţiilor de corelaţie peste toate întârzierile posibile (de la 1 la infinit) este pur şi simplu proporţională cu numărul de coincidente când preţul viitor va fi același cu cel actual, pentru alte motive decât din întâmplare. O funcție de corelație care este zero pentru toate întârzierile posibile indică prețuri aleatoare, pe când valoarea 1 corespunde corelaţiei perfecte, cand preţul se reproduce pe el insusi; o valoare constanta a prețului la orice moment *t* va indica corelatie unu, de asemenea, constantă.

Pentru seria $u(t)$, de durată T, a evalua modul în care valoarea curenta depinde de valorile istorice înseamnă a calcula funcţia (suma) de autocorelaţie, proporțională cu produsul dintre semnalul $u(t)$ şi o replică inversată și translatată în timp a acestuia $u(\tau-t)$, adică

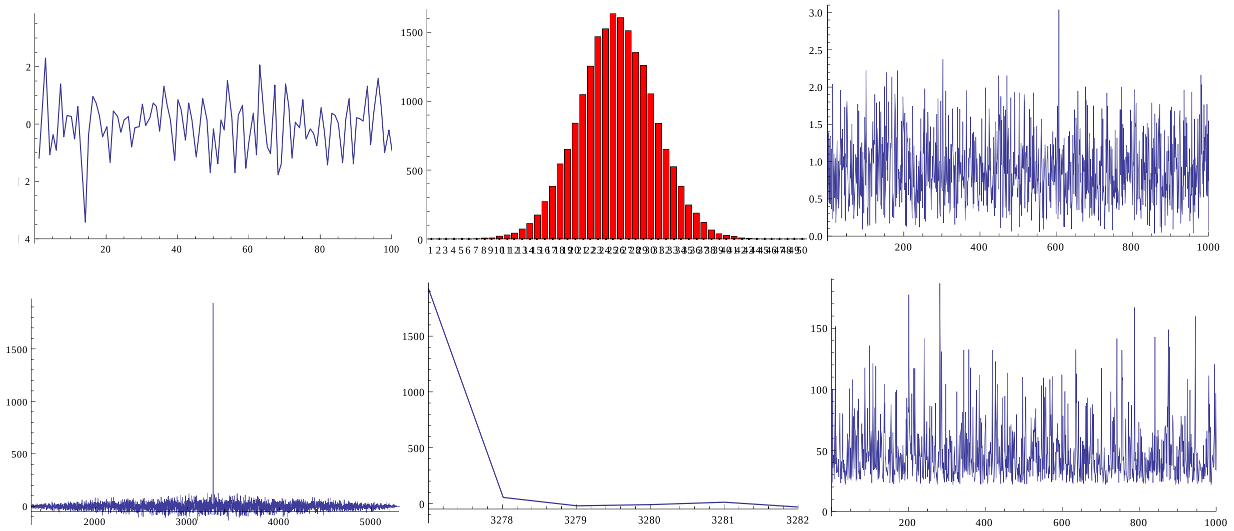
$$C_u(t) = \int_{-T}^{T} u(\tau)u(t-\tau)\mathrm{d}\tau\,; \qquad C_u(k) = \sum_{n=-N}^{n=N} u(n)u(k-n)\,.$$

Funcţia de autocorelaţie este simetrică în raport cu întârzierea $\tau(k)$, cu un maxim la $\tau=0$ ($k=0$) și, depinzând de semnalul considerat, poate avea o comportare oscilatorie, o descreştere monotonă, sau o descreştere bruscă, în cazul lipsei oricărei corelaţii.

[11] A se vedea și fascicula *Introducere*.
[12] Trading strategy to exploit correlations. P.37. D. Sornette, *Why do the markets crash*?



Cazul secvenței „drum aleator” compusă din două valori egal distribuite



Cazul secvenței compusă din valori cu distribuție Gaussiană

În exemplele de mai sus, funcția de autocorelație prezintă un maxim de tip impuls Dirac (Delta), deoarece valorile consecutive sunt necorelate, indiferent de distribuția statistică. Spectrul de frecvență este uniform, cu pantă zero, adică energia este distribuită în mod egal pe componentele spectrale.

În cazul seriilor corelate, funcția de autocorelație descrește după o lege de tip exponențial.