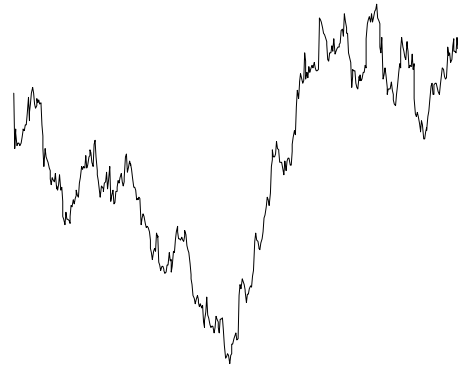




Nationalekonomiska institutionen
Ekonomihögskolan vid
Lunds Universitet
C-Uppsats

1998-06-08

Test för ickelinearitet i svenska aktieindex



Handledare:
Peter Jochumzen

Författare:
Tomas C Svensson

Abstract

The purpose of this essay is to test for nonlinearities in stock index returns.

The Brock-Dechert-Scheinkman (BDS) test of independence and identical distribution is described. By means of simulation some properties of the BDS-test are examined, when applied to time-series data generated from a fat tail probability distribution. These properties are compared with the properties of the BDS-test for simulated Gaussian time-series data. The result is that the properties of the test do not deteriorate, on the contrary, they seem to improve. The conclusion is that the BDS-test can be applied to economic time series which are known to have fat tails.

The BDS-test is then used to test for independence and identical distribution in Swedish stock index returns. Three time series are considered: Affärsvärlden's general index Jan. 1949-Dec. 1997, monthly data; Affärsvärlden's general index, May 1994-May 1998, daily data; OMX index Sept. 1994-Oct. 1994, minutely data. For all examined time series the result is that the null hypothesis of *independence and identical distribution* can be rejected at a high level of confidence, for all tested combinations of the arbitrary parameters.

After removal of possible linear dependence (autocorrelation) from the time series by means of autoregressive-moving average (ARMA) models, the BDS-test is used again, now with the null hypothesis of *no nonlinearities*. The result still is that the null can be rejected at a high level of confidence, for all time series and parameter combinations tested. The conclusion is that there are nonlinearities in stock index returns.

Innehållsförteckning

Abstract	1
Inledning	3
Bakgrund	3
Vanlig lösning på autokorrelation: ARIMA	6
Vanlig lösning på icke-konstant varians: ARCH och GARCH	6
Ett statistiskt test av oberoende och likafördelning	7
Kaos	7
Korrelationsintegralen	9
BDS-testet	10
Implementation av BDS-testet, steg för steg	11
Test av BDS-testet	15
Speciella egenskaper hos ekonomiska tidsserier	15
Cauchyfördelningen och simulering av cauchyfördelade slumpstal	16
Test av BDS-testet för cauchyfördelade slumpstal	16
Resultat och slutsats	17
Test för oberoende och likafördelning i finansiella data	18
Resultat och slutsats	19
Test för icke-linearitet i finansiella data	21
Resultat och slutsats	22
Sammanfattning	23
Källförteckning	24
Appendix I Matematiska definitioner	26
Appendix II Illustration av tjocka svansar	27
Appendix III Matlabprogram för BDS-test	29
Appendix IV SPSSkörning	31

Inledning

Syftet med denna uppsats är att testa om det finns olinjära komponenter i aktieindex. Det intressanta med olinjära komponenter och med all sorts beroende i aktieindexavkastning är att de strider mot enklare finansiella teorier. Dessa teorier gör gällande att aktieindex beskrivs fullständigt av en slumpmässig process, oberoende av den tidigare utvecklingen, en sk geometrisk Brownsk rörelse. Om ickeinjärt beroende kan påvisas innebär det att dessa teorier om stationär slumpvandring måste betraktas som en förenklad bild av verkligheten.

Ett stort antal undersökningar har gjorts av om det finns icke-linearitet och kaos i aktieindex, se Abhyankar, Copeland och Wong (1997) för en kortfattad sammanställning. Som mest har de där refererade undersökningarna använt strax under 20000 olika mätvärden som underlag, och aktieindex, valutakurser och diverse terminer har använts som tidsserier. Av de undersökningar som testat för icke-linearitet har samtliga visat på icke-linearitet.

Det som är nytt i denna uppsats, förutom att det rör sig om nya data, är att jag genom simulering undersöker om det vanligaste testet för icke-linearitet kan misstolka oberoende slumpantal från en extrem sannolikhetsfördelning, och därvid ge ifrån sig resultat som antyder att dessa slumpantal skulle vara beroende. En stegvis beskrivning ges av det ovan nämnda icke-linearitetstestet, och dessutom finns en (förhållandevis lättolkad) datorimplementation som Appendix.

Bakgrund

Vid statistiska undersökningar av ekonomiska tidsserier är det inte ovanligt, att den process f som beskriver utvecklingen y som funktion av ett antal förklarande variabler x och ev slump u , $y = f(x, u)$, är delvis okänd. Dvs tidsseriens värde y vid en viss tidpunkt kan bero av ett okänt antal förklarande variabler (ev med olika tidsförskjutning) och slumpen. Detta samband är givetvis alldeles för allmänt för att vara användbart. För att kunna bättre förstå och ev även kunna förutsäga en tidsseries utveckling är det normala att en modell ställs upp, där de variabler som man gissar bäst kan förklara y tas med. Därefter anpassas modellens parametrar till tidsserien, och med hjälp av hypotestest avgörs huruvida variablerna är signifikanta eller ej. För att få en någorlunda vettig modell tas sedan sådana variabler som inte visar sig vara signifikanta bort.

Då innehåller alltså modellen enbart signifikanta förklarande variabler, men det finns inget som säger att alla signifikanta variabler finns med. För att testa om så är fallet brukar residualerna, dvs avvikelserna för mätdata från modellens skattning $\varepsilon_t = y_t - \hat{y}_t$, undersökas närmare.

Det går att komma en bit genom att helt enkelt plotta residualerna och titta på dem. Finns det några tydliga samband är modellen ofullständig. Detta är emellertid en subjektiv metod, och den ger inget kvantitativt stöd för modellens giltighet.

De krav som ställs på residualerna från en tillräcklig modell är följande:

1. De har väntevärdet noll: $E(\varepsilon_t) = 0$.
2. De har konstant varians: $Var(\varepsilon_t) = k$, där k är en konstant.
3. De är oberoende: $p(\varepsilon_t = e | \varepsilon_{t-1}, \dots, \varepsilon_{t-\infty}) = p(\varepsilon_t = e)$, dvs sannolikheten att $\varepsilon_t = e$ påverkas inte av att man vet vad alla tidigare residualer varit.
4. De är likafördelade.

Punkterna 3 och 4 tillsammans motsvarar det som brukar förkortas IID: Independent, Identically Distributed. Punkt 2 följer egentligen av punkt 4, men brukar ändå tas upp som en egen punkt eftersom förkastning av punkt 2 är en vanlig orsak till förkastning av punkt 4.

Punkt 1 kräver att väntevärdet av residualerna är noll. Är väntevärdet inte noll beror det på att en konstant term saknas. Denna typ av fel är därför lättast att åtgärda.

För att undersöka om residualerna har konstant varians finns ett antal olika tester, t ex White's, Goldfeld-Quandt, Breusch-Pagan och Glesjer's test. Visar det sig att residualerna är heteroskedastiska (=har icke-konstant varians), och det inte finns någon självklar modifiering av den ursprungliga modellen, kan modellen i vissa fall ändras genom att lägga till tidsförskjutna skattade varianser σ_{t-s}^2 och tidsförskjutna kvadrerade residualer ε_{t-s}^2 som förklarande variabler. Detta beskrivs kortfattat i ett senare avsnitt (ARCH och GARCH).

Att undersöka om residualerna är oberoende är svårt. Det får då nämligen inte finnas något slags beroende mellan dem, varken linjärt eller ickelinjärt. Något som är lättare att testa är om de är korrelerade. Visar det sig att det finns signifikant autokorrelation så är residualerna inte oberoende. Omvändningen gäller emellertid inte; residualerna kan vara beroende även om det inte går att hitta någon signifikant autokorrelation.

Om man misstänker att residualerna inte är oberoende kan beroendet kanske (om man har tur) ha orsakats av autokorrelation, dvs ε_t kan vara signifikant korrelerad med ε_{t-s} för något $s > 0$. Detta innebär alltså att det kan finnas ett linjärt samband mellan ε_t och ε_{t-s} . För att undersöka detta finns ett par olika tester, t ex Durbin-Watson, Breusch-Godfrey, Box-Pierce och Box-Ljung. Den åtgärd som vidtas om autokorrelation kan påvisas är att inkludera tidsfördröjda residualer ε_{t-s_1} och tidsfördröjda observerade värden y_{t-s_2} i modellen.

Det finns test som testar oberoende, men tyvärr kräver dessa förhållandevis långa tidsserier. Denna uppsats kommer att beskriva ett, BDS-testet (som kräver ca 500 mätvärden). Lösningen på ett allmänt beroende mellan residualerna är emellertid sällan uppenbar.

En undersökning av om residualerna uppfyller kravet på likafördelning är inte heller det särskilt lätt att göra, men BDS-testet kan användas till att testa även detta för långa tidsserier.

Vissa av de ovan nämnda testerna är inte idealiska för att undersöka det man ofta vill undersöka när testerna används. Durbin-Watson testet används t ex på grund av sin enkelhet gärna för att testa autokorrelation, men testet signalerar ibland också om någon vanlig förklarande variabel glömts bort, varför en signal från testet inte alltid betyder autokorrelation. Som nämnts ovan kan Breusch-Godfrey testet ge indikation både vid autokorrelation och vid heteroskedasticitet, varför resultatet inte blir entydigt. BDS-testet har samma typ av problem: en signal från testet kan ha orsakats av flera olika typer av dataproblem. BDS-testet kan i vilket fall på grund av sin möjlighet att upptäcka diverse problem med residualerna också användas för att ge en viss legitimitet åt en uppställd modell, då testet kan ge en antydning till om det finns någonting signifikant kvar att förklara i residualerna eller ej.

En tidsserie bestående av residualer, för vilken punkt 1 och 2 är uppfyllda, och dessutom kovariansen $Cov(\varepsilon_t, \varepsilon_{t-s})$ bara beror av s (dvs inte av tidpunkten t), är *svagt stationär*. Detta innebär att eventuell autokorrelation inte ändras med tiden för en svagt stationär tidsserie. Om en tidsserie innehåller t ex en trend, så är tidsserien inte stationär.

En tidsserie bestående av residualer är *strikt stationär* om punkt 1, 2 och 4 är uppfyllda, och dessutom eventuellt beroende (linjärt och icke-linjärt) inte ändras med tiden. En strikt stationär tidsserie är alltså också svagt stationär. Svag stationaritet är önskvärt om t ex ARMA ska ge bra resultat.

Vanlig lösning på autokorrelation: ARIMA

Visar det sig att det finns autokorrelation, dvs ett linjärt beroende mellan residualerna, kan detta åtgärdas t ex genom att anpassa en autoregressiv modell: $AR(p)$. p stycken fördröjda tidigare värden finns då med som förklarande variabler: $y_t = a_1y_{t-1} + a_2y_{t-2} + \dots + a_p y_{t-p} + u_t$. Koefficienterna a_1, a_2, \dots, a_p kan skattas med en variant av generaliserad linjär regression, alternativt kan maximum likelihood-metoden tillgripas.

Det händer att p måste väljas mycket stort för att ta bort all signifikant autokorrelation. Det kan då löna sig att istället försöka med en glidande medelvärdesmodell: $MA(q)$, där q stycken fördröjda tidigare innovationer u är förklarande variabler:¹

$y_t = u_t + c_1u_{t-1} + c_2u_{t-2} + \dots + c_p u_{t-p}$. Koefficienterna c_1, \dots, c_p kan skattas på liknande sätt som ovan.

En kombination av de båda modellerna ovan är $ARMA(p,q)$, där både tidigare innovationer och tidigare värden tas med som förklarande variabler:

$$y_t = u_t + c_1u_{t-1} + \dots + c_p u_{t-p} + a_1y_{t-1} + \dots + a_p y_{t-p}.$$

Om tidsserien är ickestationär kan den ibland göras stationär genom differentiering. Blir resultatet av denna differentiering stationärt har man en $ARIMA(p,d,q)$ modell, där d anger differentieringens ordning.²

Vanlig lösning på icke-konstant varians: ARCH och GARCH³

Om det visat sig att variansen för residualerna inte är konstant kan detta ha orsakats av slumpmässig variation, beroende av externa faktorer, eller beroende av historiska varianser. Om det finns ett enkelt beroende av historiska varianser kan detta i vissa fall avhjälpas med t ex en ARCH-modell⁴: $\varepsilon_t = u_t \cdot \sqrt{a_0 + a_1\varepsilon_{t-1}^2 + \dots + a_p\varepsilon_{t-p}^2}$, där u_t är oberoende standardiserat normalfördelade slumpstal, a_0, \dots, a_p är värden som måste skattas, och $\varepsilon_{t-1}, \dots, \varepsilon_{t-p}$ är förklarande variabler. Detta kan också skrivas som $\varepsilon_t | \varepsilon_{t-1}, \dots, \varepsilon_{t-p} \sim N(0, \sigma_t^2)$, där $\sigma_t^2 = a_0 + a_1\varepsilon_{t-1}^2 + \dots + a_p\varepsilon_{t-p}^2$. Eftersom residualerna förutsätts ha väntevärde noll blir väntevärdet av residualerna i kvadrat $E(\varepsilon_t^2)$ lika med variansen.⁵ Därför kallas denna modell

¹ MA står för moving average.

² I:et i ARIMA står för integrated.

³ Framställningen följer Greene (1997), avsnitt 12.7 och 18.2.

⁴ ARCH=AutoRegressive Conditionally Heteroscedastic.

⁵ Jämför med omskrivningen $Var(a) = E(a^2) - (E(a))^2$.

autoregressiv med betingat beroende i varianserna (betingad heteroskedasticitet). Modellen för σ_t^2 kan tolkas som en MA(q)-modell, med fördröjda residualkvadrater istället för fördröjda funktionsvärden. För att skatta de olika koefficienterna α kan t ex maximum likelihood-metoden användas; vanlig linjär regression är olämplig eftersom de förklarande variablerna inte ingår linjärt i sambandet för ε_t .

Ett alternativ till ARCH är generaliserad ARCH (= GARCH). Generaliseringen består i att förutom kvadrerade residualer nu även betingade varianser tas med som förklarande variabler:

$\varepsilon_t | \varepsilon_{t-1}, \dots, \varepsilon_{t-q}, \sigma_{t-1}^2, \dots, \sigma_{t-p}^2 \sim N(0, \sigma_t^2)$, där nu σ_t^2 ges av

$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \dots + \alpha_q \varepsilon_{t-q}^2 + \delta_1 \sigma_{t-1}^2 + \dots + \delta_p \sigma_{t-p}^2$. Här är även $\delta_1, \dots, \delta_p$ koefficienter som

måste skattas, och $\sigma_{t-1}^2, \dots, \sigma_{t-p}^2$ är nya förklarande variabler. Denna modell kan tolkas som en ARMA(p,q)-modell, med residualkvadrater istället för fördröjda funktionsvärden och med fördröjda betingade residualvarianser istället för fördröjda residualer. Skattning av koefficienterna kan även här göras med maximum likelihood-metoden; linjär regression bör undvikas av samma anledning som ovan.

En ARCH-modell ger ibland fler signifikanta koefficienter (dvs fler tidsfördröjda variabler måste tas med) än en GARCH-modell, varför ARCH-modeller kan se mer komplicerade ut. För GARCH-modeller är det å andra sidan betydligt svårare att skatta koefficienterna och testa deras signifikans. Därför är en ARCH-modell med ett fåtal koefficienter oftast att föredra om det är möjligt att välja.

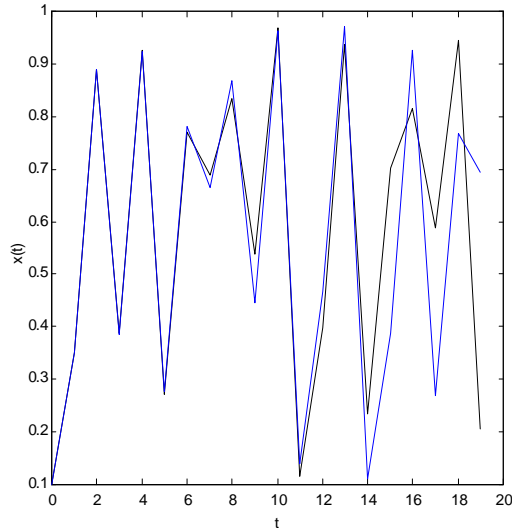
Ett statistiskt test av oberoende och likafördelning

Antalet statistiska tester som har IID (oberoende och likafördelning) som nollhypotes är, som nämnts mycket begränsat. Ett av dessa tester är BDS-testet, som har sitt ursprung i teorin för kaos.

Kaos

Kaosteorin behandlar dynamiska system vars beteende är i hög grad deterministiskt (dvs styrs av exakta lagar), men som ändå på grund av systemets konstruktion är praktiskt taget omöjligt att förutspå bortom ett begränsat antal steg. Ett välkänt exempel på ett sådant system är väder, som bara kan förutsägas ett par dagar framåt. Ett av de viktigaste kännetecknen på kaos är ett systems känslighet för begynnelsevärden; detta förkortas ofta SIC = Sensitivity to

Initial Conditions. Detta innebär att en liten ändring av begynnelsevärdena gör att systemet efter ett par perioder utvecklar sig väldigt annorlunda jämfört med om ingen sådan ändring av begynnelsevärdena hade gjorts.



Figuren här intill illustrerar som exempel SIC för den ganska enkla funktionen $x_t = 3,9 \cdot (x_{t-1} - x_{t-1}^2)$. Den svarta kurvan hade begynnelsevärdet $x_0 = 0,1000$ och den gråa $x_0 = 0,1002$. Under de första stegen fram till $t \approx 7$ märks ingen större skillnad. Därefter avviker de två kurvorna allt mer från varandra.

Att funktionen i exemplet är icke-linjär är ingen tillfällighet. Ett nödvändigt villkor för att kaos ska kunna uppkomma är att det finns något icke-linjärt beroende.

För att beskriva ett systems utveckling sätts de n stycken variabler, som anger systemets tillstånd vid en viss tidpunkt, i en (n -dimensionell) vektor. Den punkt som vektorn pekar ut sätts sedan in i ett (n -dimensionellt) tillståndsrum. När tiden går och systemet ändrar tillstånd flyttar sig denna punkt i (det n -dimensionella) tillståndsrummet. Den kurva som bildas då punkten flyttar sig kallas en *trajektoria*. För en kaotisk tidsserie går denna trajektoria ofta inte hur som helst i tillståndsrummet, utan den bildar ett mönster, precis som om den attraherades av en viss kurva. Den kurva som den attraheras av kallas *säregen attraktor*. Dimensionen hos den säregna attraktorn behöver inte vara ett heltal.⁶

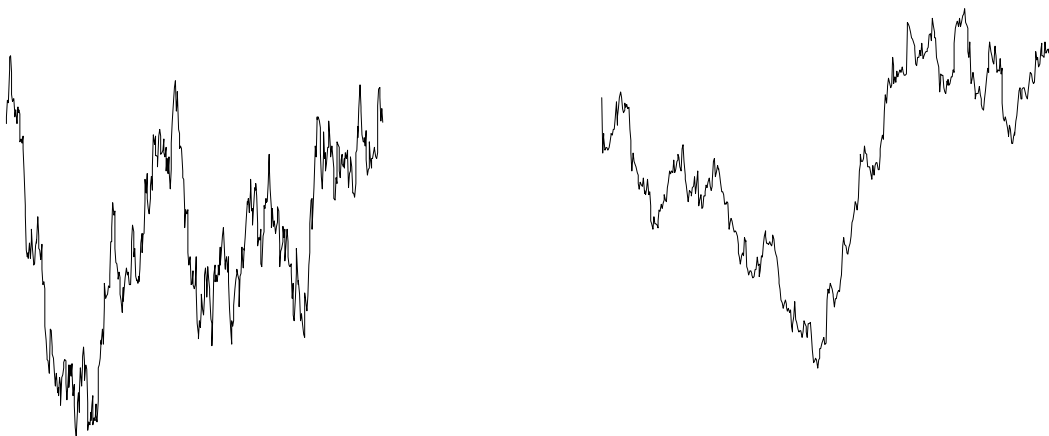
För systemet i exemplet ovan är antalet tillståndsvariabler $n = 1$, nämligen tillståndsvariabeln x_t . Det n -dimensionella tillståndsrummet blir i detta fall 1-dimensionellt, dvs en linje, och denna linje är någon valfri linje (i exemplet parallell med den vertikala axeln i figuren). Trajektorian för två olika begynnelsestillstånd är plottad mot en tidsaxel i figuren ovan.

I praktiken är tyvärr tillståndet x oftast okänt, varför det måste återskapas. Detta kan göras genom att helt enkelt ur en tidsserie ta ett antal (m stycken) punkter i tidsföljd och lägga dem i en vektor. Om m är tillräckligt stort i förhållande till n kan sedan antas att den nya

⁶ En kurva kan ha en dimension >1 om den på något sätt 'fyller upp' en större del av den tillgängliga rymden än vad t ex en linje eller cirkel gör.

(m -dimensionella) vektorn beskriver kanske inte samma, men åtminstone ett liknande tillstånd som (den n -dimensionella) tillståndsvektorn.⁷

Att med ögat skilja en tidsserie med ursprung i ett kaotiskt system från en tidsserie med ursprung i ett ekonomiskt system kan vara svårt. Figurerna nedan visar dels OMXindex minut för minut, och dels en kaotisk funktion (Weierstrass funktion). Det bör inte vara uppenbart vilken som är vilken.⁸ Denna likhet mellan en deterministisk och en (troligen) stokastisk process är en av orsakerna till det intresse som visats för metoder som används till att beskriva kaotiska system, och om de även kan användas på ekonomiska system.



Korrelationsintegralen

Ett begrepp som visat sig användbart inom ekonometri är korrelationsintegralen, som betecknas $C_{m\epsilon}$. Den anger den ungefärliga sannolikheten att en slumpvist utvald punkt på trajektorian i m -rummet är 'nära' en annan slumpvist utvald punkt på trajektorian. (Hur nära som ska anses vara 'nära' bestäms av parametern ϵ , vilken ofta anges i termer av standardavvikelse.) Korrelationsintegralen räknar nämligen antalet par av punkter som är nära varandra och dividerar med antalet möjliga val av (olika) punktpar.

Korrelationsdimensionen d är gränsvärdet av $d_m = \frac{\log C_{m\epsilon}}{\log \epsilon}$ då m växer samtidigt som ϵ blir litet. Om det visar sig att det inte finns något gränsvärde, dvs d_m ökar när m ökar, innebär det

⁷ Anledningen till att man inte behöver ta med hur många gamla värden som helst är att sådana på grund av SIC så snabbt förlorat sin inverkan på det tillstånd som ska skattas.

⁸ Den vänstra figuren är Weierstrass funktion, och den högra är OMXindex.

att processen har en oändligdimensionell attraktor, vilket är detsamma som att den är stokastisk. På grund av att antalet mätpunkter i praktiken alltid är begränsat går det emellertid inte att skilja mellan en attraktor med hög dimension och en med oändlig dimension. Ett annat problem med korrelationsdimensionen är att den inte ger någon information om signifikansnivån för resultatet.

Vad ekonomiska tidsserier beträffar har det hittills varit svårt att finna belägg för kaos, kanske beroende på att de undersökta tidsserierna har varit för korta, eller därför att kaos eventuellt inte kan upptäckas över bruset. Alternativt kanske där inte finns något kaotiskt att upptäcka.

BDS-testet

BDS-testet, som utvecklats av Brock, Dechert och Scheinkman, utgår från korrelationsintegralen, men förutsätter ingenting om huruvida kaos förekommer i den undersökta tidsserien eller ej. Utan att göra några antaganden om typ av fördelning kan det visas att för oberoende och likafördelade mätvärden gäller $C_{m\epsilon} = (C_{1\epsilon})^m$, där likhetstecknet för ändliga dataserier måste tolkas statistiskt. Däremot är det sällsynt att likheten är uppfylld (utom i specialkonstruerade fall) för mätvärden som inte är oberoende eller likafördelade. Skillnaden $C_{m\epsilon} - (C_{1\epsilon})^m$ ger således ett mått på om mätvärdena är oberoende och likafördelade eller ej. För att kunna utföra ett statistiskt test av detta måste fördelningen för skillnaden vara känd. Det visades av Brock, Dechert och Scheinkman att fördelningen under nollhypotesen om IID för långa mätserier är $N(0, \frac{V_{m\epsilon}}{\sqrt{L}})$, där L anger antalet mätpunkter och $V_{m\epsilon}$ är ett komplicerat uttryck (se nedan).

Genom att jämföra det standardiserade värdet $W_{m\epsilon} = \sqrt{L} \frac{C_{m\epsilon} - (C_{1\epsilon})^m}{V_{m\epsilon}}$ med någon lämplig normalfördelningskvantil kan alltså hypotesen om oberoende och likafördelning antingen förkastas eller inte förkastas. Uttrycket ska vid IID vara fördelat enligt $N(0,1)$, dvs $W_{m\epsilon} \sim N(0,1)$. $W_{m\epsilon}$ kan således helt enkelt jämföras med ± 1.96 för 95% konfidens, eller ± 2.58 för 99%, precis som vanligt. Hypotestest är sedan lätt att göra:

nollhypotesen är H_0 : residualerna är IID,
och mothypotesen H_1 : residualerna är inte IID.

Nollhypotesen kan förkastas om $|W_{m\epsilon}| > 1.96$ resp $|W_{m\epsilon}| > 2.58$.

BDS-testet har enligt Brock, Hsieh och LeBaron (1993) inte bra egenskaper för dataserier kortare än 500 datapunkter. Kravet att få ihop mätserier med minst 500 mätpunkter är i

ekonomisk analys en allvarlig begränsning. När det gäller t ex BNP eller arbetslöshet går det kanske att få tag på mätvärden för varje månad. Mätperioden skulle då behöva sträcka sig över $\frac{500}{12} \approx 40$ år. Gäller undersökningen istället finansiella data blir det lättare, då mätvärden går att få med ett par sekunders intervall. Detta gör att ett par timmar räcker för att få ihop 500 finansiella mätvärden.

Implementation av BDS-testet, steg för steg

Residualerna χ_i bör ges den genomsnittliga standardavvikelsen 1.⁹ Detta är det normala i litteraturen, och det görs främst för att slippa definiera ε som $\varepsilon = k\sigma$. Istället sätts alltså standardavvikelsen $\sigma = 1$, så att $\varepsilon = k$, där $k > 0$ är en valbar konstant. För att residualerna ska få variansen 1 beräknas $x_i = \frac{\chi_i}{\sigma_\chi}$, där σ_χ är den skattade standardavvikelsen för χ , och x_i är residualer standardiserade till variansen (och standardavvikelsen) 1.

Välj nu ε på något lämpligt sätt. Definitionen av korrelationsintegralen kräver för att ge så bra resultat som möjligt att ε väljs litet. ε kan emellertid inte väljas hur litet som helst för ändliga tidsserier, eftersom då inga punkter ligger 'nära' varandra. Det vanligaste är att välja något värde mellan 0.5 och 1.5 standardavvikelser. Eftersom vi nu har standardavvikelsen 1 väljs ε i intervallet $[0.5; 1.5]$.¹⁰ Ett ε ur detta intervall brukar ge lagom många punkter som är 'nära' varandra i det m -dimensionella rummet. Detta m måste också väljas av den som gör undersökningen, och det normala är att $m \in [2; 10]$.¹¹ Eftersom man sällan vet vad m borde vara är det vanligt att beräkna testet för flera olika m . Den beskrivning som ges här kräver att m är ett heltal. Denna inskränkning är varken ovanlig eller orimlig.

Det som följer nedan är med nödvändighet avsevärt mera tekniskt och matematiskt detaljerat. Det Matlab-program som visas i Appendix III är en implementation av BDS-testet enligt beskrivningen nedan. θ -funktioner kommer att användas för att räkna antalet punkter som är 'nära' varandra, se gärna Appendix I för matematiska definitioner. Sättet att mäta närhet mellan punkter i ett m -dimensionellt rum bestäms av vilken norm som används, se Appendix I för en definitionen av norm (betecknas $\|\cdot\|$). Likheten $C_{m\varepsilon} = (C_{1\varepsilon})^m$ som BDS-testet bygger på tycks kräva att max-normen används. Detta är dessutom den norm som kräver minst räknearbete.

⁹ Standardavvikelse för mätvärden går alltid att beräkna.

¹⁰ Detta beteckningssätt utläses: slutna intervallet från 0.5 till 1.5. Ett *slutet* intervall innebär att även gränserna tillhör intervallet.

¹¹ Detta utläses: m ligger i slutna intervallet mellan 2 och 10.

Låt L vara antalet mätpunkter, dvs antalet x_i . Definiera $x_{im} = (x_i, x_{i+1}, \dots, x_{i+m})$, dvs vektorer bestående av m i tiden efter varandra följande observationer x_i ; antalet sådana vektorer som kan skapas är $L-m+1$. Dessa x_{im} är de tidigare nämnda m -vektorer som ska approximera tillståndet för en kaotisk process. Då kan korrelationsintegralen $C_{m\epsilon}$ approximeras:

$$C_{m\epsilon} = \binom{L-m+1}{2}^{-1} \sum_{s=1}^{L-m} \sum_{t=s+1}^{L-m+1} \theta(\epsilon - \|x_{sm} - x_{tm}\|). \quad \text{Detta är som synes ingen egentlig integral.}$$

Uttrycket ovan blir emellertid samma som korrelationsintegralen i gränsen $n \rightarrow \infty$ och $\epsilon \rightarrow 0$. Uttrycket kan kanske te sig något komplicerat. En uttydning av beståndsdelarna ges därför: $\|x_{sm} - x_{tm}\|$ kan sägas beteckna avståndet mellan de båda vektorerna x_{sm} och x_{tm} (punkter i det m -dimensionella rummet). Om avståndet är mindre än ϵ blir θ -funktionen 1, annars blir den 0. Summorna räknar sedan antalet punkter som är tillräckligt nära varandra, dvs antalet gånger θ -funktionen blir 1.

Att summationsgränserna ser ut som de gör beror dels på att det räcker att ta med varje par av punkter en gång, eftersom avstånden $\|x_{sm} - x_{tm}\|$ och $\|x_{tm} - x_{sm}\|$ är lika stora (det spelar ingen roll om avståndet mäts från punkten x_{sm} till punkten x_{tm} , eller från punkten x_{tm} till x_{sm} , avståndet blir lika stort i båda riktningarna), och dels på att det är ointressant att ta med $\|x_{sm} - x_{sm}\|$ eftersom det där inte är två olika punkter (avståndet mellan en punkt x_{sm} och sig själv är 0).

Divisionen med binomialkoefficienten $\binom{L-m+1}{2}$ är sedan ett slags standardisering, så att $C_{m\epsilon}$ ska kunna tolkas som en sannolikhet: Om alla möjliga par av punkter ligger nära varandra blir $C_{m\epsilon} = 1$, och om alla är icke-nära blir $C_{m\epsilon} = 0$.

Om även $C_{1\epsilon}$ beräknas, dvs korrelationsintegralen med $m=1$, kan nu differensen $T_{m\epsilon} = C_{m\epsilon} - (C_{1\epsilon})^m$ beräknas. Differensen $T_{m\epsilon}$ blir = 0 om den beräknas för en serie bestående av oändligt många residualer som är IID, och den blir $\neq 0$ om residualerna inte är IID. I verkligheten har man givetvis aldrig oändligt många mätvärden, och därför blir skillnaden i allmänhet inte exakt 0 även om mätvärdena är IID. För att kunna veta om avvikelser från noll är signifikant eller inte måste nu differensen $T_{m\epsilon}$ standardiseras. Som nämnts tidigare är fördelningen för $T_{m\epsilon}$ vid oberoende och likafördelning $N(0, \frac{V_{m\epsilon}}{\sqrt{L}})$. Eftersom vi vet att den har väntevärdet 0 behöver vi nu 'bara' beräkna dess varians.

Den asymptotiska variansen ges enligt Brock et al (1996) av:

$$V_{m\epsilon}^2 = 4m(m-2)C_{\epsilon}^{2m-2}(K_{\epsilon} - C_{\epsilon}^2) + K_{\epsilon}^m - C_{\epsilon}^{2m} + 2 \sum_{j=1}^{m-1} (C_{\epsilon}^{2j}(K_{\epsilon}^{m-j} - C_{\epsilon}^{2m-2j}) - mC_{\epsilon}^{2m-2}(K_{\epsilon} - C_{\epsilon}^2)).$$

Den intresserade kan läsa om varför denna ser ut precis som den gör in *ibid*. Observera att

C_ε , med bara ett index, inte är detsamma som korrelationsintegralen $C_{m\varepsilon}$, även om den beräknas på liknande sätt: $C_\varepsilon = \frac{1}{L^2} \sum_{s=1}^L \sum_{t=1}^L \theta(\varepsilon - \|x_{s1} - x_{t1}\|)$. Normen beräknas alltså i detta fall för skillnaden mellan två skalärer x_{s1} och x_{t1} . Tolkningen är liknande den tidigare: C_ε är andelen par av punkter som är nära varandra, men här tas även hänsyn till ordning, och punkterna behöver inte heller vara olika. Vid datorimplementation skulle symmetrin kunna utnyttjas och en del beräkningstid kunna sparas genom att låta s istället variera från 1 till $L-1$, och t gå från $s+1$ till L , med insikt om att varje par av punkter förekommer två gånger, och dessutom avståndet mellan en punkt och sig själv alltid är mindre än ε .

K_ε påminner i sin tur om C_ε , och definieras:

$K_\varepsilon = \frac{1}{L^3} \sum_{r=1}^L \sum_{s=1}^L \sum_{t=1}^L \theta(\varepsilon - \|x_{r1} - x_{s1}\|) \theta(\varepsilon - \|x_{s1} - x_{t1}\|)$. Tolkningen är att K_ε är andelen tripplar av tal, sådana att x_{r1} är nära x_{s1} samtidigt som x_{s1} är nära x_{t1} .

Beräkningen av K_ε kan ställa till problem om den implementeras oförsiktigt. Skrivs den som tre stycken nästlade loopar, blir komplexiteten $O(L^3)$, vilket utläses att beräkningstiden växer som antalet mätvärden upphöjt till 3. Det är heller inte tillräckligt att ta hänsyn till liknande symmetriegenskaper som de som finns hos C_ε , eller att skriva programmet i ett lågnivåspråk. Symmetriegenskaperna sparar in uppskattningsvis 7/8 av de ursprungliga beräkningarna. Antag dessutom att lågnivåspråket (för att ge enkla överslagsräkningar) blir 64 gånger snabbare än det interpreterande Matlab. Då kan ändå 'bara' $(8 \cdot 64)^{\frac{1}{3}} = 8$ gånger mer data bearbetas under en given tidsrymd. Bättre är då att istället skriva om beräkningen av K_ε så att den blir $O(L^2)$:¹²

Det framgår att det för en given mängd data x_i finns ett begränsat (dvs inte oändligt) antal argument till θ -funktionen som är intressanta. Börja med att beräkna θ -funktionens värden för alla dessa och kalla dem τ_{rs} : $\tau_{rs} = \theta(\varepsilon - \|x_{r1} - x_{s1}\|)$; detta ger L^2 stycken funktionsvärden, som antingen är 0 eller 1. Anordna dessa i en matris (matrisen blir symmetrisk, och alla diagonalelement blir 1):

$$\begin{bmatrix} \tau_{11} & \tau_{12} & \dots & \tau_{1L} \\ \tau_{21} & \tau_{22} & \dots & \tau_{2L} \\ \dots & \dots & \dots & \dots \\ \tau_{L1} & \tau_{L2} & \dots & \tau_{LL} \end{bmatrix}. \quad (1)$$

¹² Efter att ha gjort denna omskrivning av algoritmen, lönar det sig givetvis att utnyttja symmetrin och att använda lågnivåspråk. Matlab-programmet i Appendix III utnyttjar inte symmetrin fullt ut, då detta skulle försvåra utnyttjandet av matrisräkning, och därför inte ger någon nämnvärd tidsvinst i Matlab.

Beräkna sedan summan $\Theta_i = \sum_{j=1}^L \tau_{ij}$ av varje rad¹³ i matrisen (1), och lagra dessa i en vektor Θ .

Lägg därefter samman skalärprodukten av varje rad i matrisen (1) med vektorn Θ , dvs

beräkna $\kappa = \sum_{i=1}^L \left(\begin{bmatrix} \tau_{i1} & \tau_{i2} & \dots & \tau_{iL} \end{bmatrix} \cdot \Theta \right)$. Detta κ (kappa) är värdet av trippelsumman, och

$$K_\varepsilon = \frac{\kappa}{n^3}.$$

Att κ verkligen blir lika med $\sum_{r=1}^L \sum_{s=1}^L \sum_{t=1}^L \theta(\varepsilon - \|x_{r1} - x_{s1}\|) \theta(\varepsilon - \|x_{s1} - x_{t1}\|)$ ska nu visas.

Detta stycke kan därför överhoppas utan att sammanhanget går förlorat.

Låt som ovan $\tau_{rs} = \theta(\varepsilon - \|x_{r1} - x_{s1}\|)$. Att $\tau_{rs} = \tau_{sr}$ och $\tau_{rr} = 1$ framgår direkt ur denna definition. Uttrycket för trippelsumman kan nu enkelt skrivas om genom att flytta τ_{rs} utanför summationen över t :

$$\sum_{r=1}^L \sum_{s=1}^L \sum_{t=1}^L \tau_{rs} \tau_{st} = \sum_{r=1}^L \sum_{s=1}^L \tau_{rs} \sum_{t=1}^L \tau_{st} = \sum_{r=1}^L \sum_{s=1}^L \tau_{rs} \Theta_s. \quad ^{14}$$

Vi ser nu att summan över s kan tolkas som en skalärprodukt av vektorerna $\tau_r = (\tau_{r1}, \tau_{r2}, \dots, \tau_{rL})$ och Θ . Alltså är

$$\sum_{r=1}^L \sum_{s=1}^L \tau_{rs} \Theta_s = \sum_{r=1}^L (\tau_r \cdot \Theta),$$

vilket avslutar härledningen.

Det egentliga testvärdet kan slutligen beräknas som $W_{me} = \frac{\sqrt{L} T_{me}}{V_{me}}$. Vid IID gäller, som nämnts, att $W_{me} \sim N(0, 1)$.

En sammanfattning av de ovanstående stegen ges i ett försök att öka överskådligheten:

1. Ta fram residualer från någon modell.
2. Standardisera dessa till variansen 1.
3. Välj ε och m .
4. Skapa de m -dimensionella vektorerna x_{sm} och de 1-dimensionella vektorerna x_{s1} .
5. Beräkna avståndet mellan alla m -dimensionella vektorer och jämför med ε . Lagra resultatet av jämförelsen (0 eller 1) i en matris (denna benämns senare $p5$).
6. Beräkna avståndet mellan alla 1-dimensionella vektorer och jämför med ε . Lagra resultatet av jämförelsen (0 eller 1) i en matris (denna benämns senare $p6$).

¹³ Eller kolumn, matrisen är ju symmetrisk.

¹⁴ Uttrycket kan förenklas ytterligare på samma sätt, så att beräkningen blir $O(L)$.

Anledningen till att detta inte fullföljs här är att andra delar av beräkningen ändå är $O(L^2)$, varför tidsvinsten blir $O(1)$. Tack till David Edgerton för denna upplysning.

7. Beräkna dessa matrises respektive radsummor. Lagra resultatet i två kolumnvektorer.
8. Beräkna kolumnsummorna i dessa vektorer var för sig.
9. Beräkna summan av skalärprodukten av varje rad i respektive matris med respektive vektor.
10. Dividera resultatet från beräkningen på $p6$ i punkt 9 med L^2 . Resultatet blir då C_ε .
11. Dividera resultatet från beräkningen på $p6$ i punkt 9 med L^3 . Resultatet blir då K_ε .
12. Beräkna summan av elementen ovanför diagonalen i matriserna $p5$ och $p6$ var för sig.
13. Dividera resultatet av beräkningen på $p5$ i punkt 12 med $\frac{1}{2}(m-L)(m-L-1)$. Resultatet blir $C_{m\varepsilon}$.
14. Dividera resultatet av beräkningen på $p6$ i punkt 12 med $\frac{1}{2}(L-1)L$. Resultatet blir $C_{1\varepsilon}$.
15. Beräkna $T_{m\varepsilon}$ som $T_{m\varepsilon} = C_{m\varepsilon} - (C_{1\varepsilon})^m$.
16. Beräkna variansen enligt uttrycket för $V_{m\varepsilon}$. Allt som ingår i det är beräknat ovan.
17. BDS-testvärdet beräknas slutligen som $W_{m\varepsilon} = \frac{\sqrt{L} T_{m\varepsilon}}{V_{m\varepsilon}}$.

Test av BDS-testet

Speciella egenskaper hos ekonomiska tidsserier

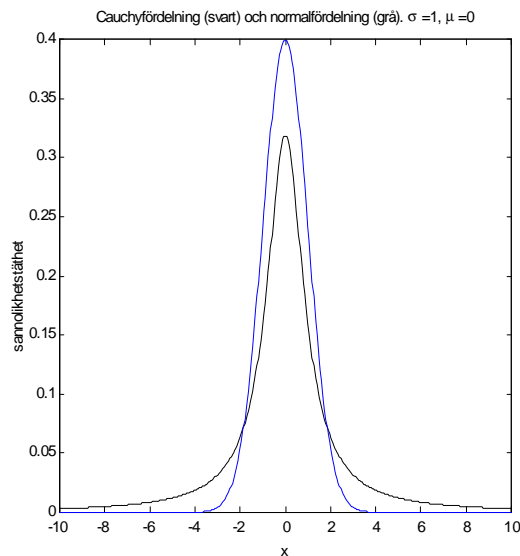
Ett vanligt antagande inom olika typer av ekonomisk teori är att residualerna till uppställda modeller är oberoende och normalfördelade. I verkliga tidsserier är förmodligen så inte alltid fallet, men ofta kan dessa modeller ändå fungera ganska bra.

Ett exempel på en sådan avvikelse från normalfördelningen antyds i en snabb liten illustration (som inte är huvudsyftet med denna uppsats): I finansiella tidsserier av typen aktieindex säger den tillämpliga teorin att värdet på lång sikt växer exponentiellt, dvs som ränta på ränta, och att tillväxten sker oberoende av tidigare tillväxt. I Appendix II visas olika histogram över indextillväxt.¹⁵ I varje histogram visas andelen indextillväxter som faller inom varje intervall. Arean under kurvan i ett intervall anger sannolikheten att en indextillväxt hamnar inom intervallet. På x-axeln är avsatt intervallens (logaritmerade) gränser. Av diagrammen bör framgå att normalfördelningen stämmer dåligt med dessa mätvärden. Det syns tydligast på de med logaritmerad y-axel (till höger) att histogrammens svansar, dvs de yttre delarna, är något tjockare än normalfördelningens svansar.

¹⁵ Med indextillväxt menar jag $\frac{\text{värdet av aktieindex vid tiden } t}{\text{värdet av aktieindex vid tiden precis före } t}$.

Cauchyfördelningen och simulering av cauchyfördelade slumpstal

En fördelning som är bekant för att ha extremt tjocka svansar är cauchyfördelningen, även känd som lorentzfördelningen. Förutom denna egenskap görs i föreliggande uppsats inga uttalanden om andra likheter eller övriga gemensamma egenskaper med verkliga ekonomiska fördelningar. Täthetsfunktionen för cauchyfördelningen är $f(x) = \frac{\sigma^2}{\pi(\sigma^2+(x-\mu)^2)}$, och fördelningsfunktionen är $F(x) = \frac{\sigma}{\pi} \arctan(\frac{x-\mu}{\sigma})$.



Diagrammet här intill visar sannolikhetstätheten för cauchyfördelningen (svart kurva) tillsammans med normalfördelningen (grå kurva).

Utgående från rektangelfördelade slumpstal¹⁶ u , kan slumpstal x med (godtycklig) fördelningsfunktion F fås med hjälp av inversmetoden $x = F^{-1}(u)$. För cauchyfördelningen innebär detta $x = \mu + \sigma \cdot \tan(\frac{\pi u}{\sigma})$. Tal som är approximativt rektangelfördelade fås från en vanlig

slumptalsalgoritm, dvs 'slumptalen' räknas fram av datorn. Detta gör att 'slumptalen' egentligen inte är ett dugg slumpmässiga, utan är fullständigt deterministiska. Algoritmens komplexitet gör dock att för alla praktiska tillämpningar kan slumptalen (eller pseudo-slumptalen som de också kallas) ändå antas vara oberoende. En slumptalsalgoritm är för övrigt ett exempel på en kaotisk process med en så hög korrelationsdimension d , att denna i praktiken inte går att skilja från en process med oändlig korrelationsdimension, dvs en stokastisk process.

Test av BDS-testet för cauchyfördelade slumpstal

Det går att bevisa teoretiskt att BDS-testet fungerar, oberoende av vilken sorts fördelning talen som undersöks har. Emellertid är det inte alls självklart att det fungerar lika bra för korta tidserier med tjocka svansar.

¹⁶ Rektangelfördelade slumpstal har lika stor sannolikhet att anta alla värden inom ett intervall, i detta fall intervallet $[0,1]$. Om täthetsfunktionen ritas i diagram har den formen av en rektangel.

För att testa om BDS-testet skulle kunna användas till ekonomiska tidsserier, även sådana med tjocka svansar, ska nu BDS-testets uppförande undersökas när det matas med slumpstal från den något extrema cauchyfördelningen. För att ha något välkänt att jämföra med matas BDS-testet också med slumpstal från normalfördelningen.

Eftersom BDS-testet enligt Brock, Hsieh och LeBaron (1993) inte har bra egenskaper för dataserier kortare än 500 datapunkter, och av hänsyn till tillgänglig beräkningskapacitet används här dataserier med 500 slumpstal i varje. På grund av de något osäkra (och mer eller mindre godtyckliga) parametrarna ε och m till testet prövades fyra olika kombinationer, där $\varepsilon = 0.5$ resp 0.8 , och $m = 2$ resp 4 . För varje sådan kombination av ε och m , samt för både normal- och cauchyfördelningen beräknas testvärdet 200 gånger, vilket bör vara tillräckligt för att ge en uppfattning om testets egenskaper för cauchyfördelade slumpstal.¹⁷

Resultat och slutsats

Simuleringen gjordes i Matlab på 25 st samtida Sun Ultra 1 arbetsstationer (jämförbara med Pentium II 233) och tog strax över 2 timmar.

Antalet gånger $|W_{m\varepsilon}| > 1.96$, dvs antalet förkastningar på 5%-nivån av IID-hypotes enligt BDS-testet följer i tabellform:

	Cauchyfördelning		Normalfördelning	
m=	2	4	2	4
epsilon=				
0,5	10	11	11	23
0,8	9	10	12	14

Om värdena i ovanstående tabell divideras med antalet simuleringar (200 st) erhålls andelen förkastningar:

	Cauchyfördelning		Normalfördelning	
m=	2	4	2	4
epsilon=				
0,5	5,0%	5,5%	5,5%	11,5%
0,8	4,5%	5,0%	6,0%	7,0%

BDS-testet förkastar ibland nollhypotesen, precis som alla andra statistiska test, att slumpstal är IID, trots att den hypotesen är sann. På 5%-nivån är det förväntade värdet på alla tal i ovanstående andel-förkastningar-matris 5%, åtminstone om längden på de dataserier som BDS-testet körs på kan anses stor. Det framgår att resultatet för cauchyfördelningen ligger

¹⁷ Således har $2 \cdot 2 \cdot 2 \cdot 200 = 1600$ beräkningar av BDS-testvärden gjorts.

närmre de förväntade värdena än normalfördelningen. Detta är testat för flera olika kombinationer av m och ε , varför slutsatsen av undersökningen måste vara att BDS-testet faktiskt verkar fungera bättre för cauchyfördelade slumpstal än för normalfördelade, i alla fall i området kring dessa m och ε , och för 500 eller fler slumpstal. Ekonomiska tidserier bör kunna anses ligga någonstans mellan normalfördelningen och cauchyfördelningen beträffande svansarnas storlek. Denna undersökning kan därför inte förkasta BDS-testet som metod inom ekonomisk tidsserieanalys, utan tvärtom verkar alltså de tjocka svansarna ha en positiv inverkan på BDS-testets egenskaper i korta serier.

Test för oberoende och likafördelning i finansiella data

Samma finansiella data som använts för att illustrera de tjocka svansarna i ekonomiska tidsserier kommer nu att användas till att testa förekomst av icke-lineariteter i finansiella tidsserier. De tre serier som används är:

- 1a. Affärsvärldens generalindex (Afgx) jan 1949-dec 1997, månadsvis, 588 observationer.
- 2a. Affärsvärldens generalindex maj 1994-maj 1998, dagsvis, 879 observationer.
- 3a. OMX-index sept 1994-okt 1994, minutvis, 10655 observationer.¹⁸

Affärsvärldens generalindex är ett värdeviktat index av bolagen på Stockholmsbörsens A-lista.

OMX-index är ett värdeviktat index av de 30 mest omsatta bolagen på Stockholmsbörsen.

Minutdata i serie 3a kan ställa till problem genom att börshandeln inte fungerar på precis samma sätt vid börsens öppningen som under resten av dagen. En extra tidsserie 3b har därför skapats innehållande enbart avkastning hänförlig till tiden 10:15-16:00. Denna innehåller 10161 indexavkastningar.

Tidsserie 3a och 3b ställer dessutom till med tekniska problem, genom att de är förhållandevis långa. De delas därför upp i 3 ungefär lika stora delar med ca 3500 punkter i varje.¹⁹

Månadsdata i serie 1a kan också ställa till problem genom att den sträcker sig över så lång tid, under vilken börshandeln förändrats en hel del, t ex genom införandet av ett elektroniskt handelssystem.

¹⁸ Tidsserien OMX-index med minutvisa observationer tillhandahölls ursprungligen av Lars Nordén, Nationalekonomiska institutionen, Lunds universitet.

¹⁹ Så som Matlabprogrammet i Appendix III är skrivet skulle det annars behövas minnesutrymme i storleksordningen gigabyte.

Dagsdata i serie 2a ger även den upphov till en hel del problem. I tidsserien saknas nämligen vissa handelsdagar, varför avkastningar enbart beräknats för de dagar då det inte saknats något av de två värden som behövs. Man skulle också kunna misstänka att avkastning över veckosluten inte beter sig likadant som avkastning under veckan. En extra tidsserie 2b har därför skapats där avkastning över veckosluten (fre-mån) inte finns med. Tidsserie 2b består av 709 värden.

Indexvärdet antas öka exponentiellt (på lång sikt), liksom ränta-på-ränta.²⁰ Därför räknas tidsserien om till en sorts ränta, eller tillväxtfaktor: $1 + r_t = \frac{\text{indexvärde vid tiden } t}{\text{indexvärde vid tiden närmast innan } t}$.

Eftersom börsvärdet i princip kan variera mellan 0 och oändligheten ($[0, \infty)$) blir det intervall som tillväxtfaktorn $1 + r_t$ kan variera inom osymmetriskt: $[0, \infty[$. Genom att logaritmera tillväxtfaktorn blir intervallet symmetriskt: $\log(1 + r_t)$ kan då variera inom $] -\infty, \infty[$. Denna logaritmering påverkar inte den eventuella förekomsten av egenskaperna oberoende eller likafördelning, men den bör ge en förväntad fördelningsfunktion som är symmetrisk. Det test som gjordes ovan av BDS-testet utfördes med simulerade data från symmetriska fördelningar, varför ett underlåtande att göra logaritmeringen skulle kunna kompromettera den fortsatta undersökningen. Eftersom r_t också förväntas ha ett positivt medelvärde (dvs $\neq 0$) har detta subtraherats bort. Tillväxtfaktorerna minskade med medelvärdet ska därefter enligt teorin vara oberoende och likafördelade (och givetvis ha väntevärde = 0).

Vad själva testet beträffar har alla ovanstående serier fler än 500 mätpunkter, varför BDS-testet kan förväntas vara något närmare sin asymptotiska fördelning än för de ovan testade slumptalsserierna. De följande testerna kommer också att utföras för samma val av kombinationer av m och ϵ .

Resultat och slutsats

Nedanstående tabeller visar BDS-värdet $W_{m\epsilon}$ för några olika val av m och ϵ .

W	Afgx månad		Afgx dag		Afgx dag utan helger	
	2	4	2	4	2	4
epsilon=						
0.5	5,34	9,64	3,91	5,71	3,97	6,14
0.8	5,33	9,92	3,60	5,82	3,67	5,89

²⁰ Ränta på ränta enligt den vanliga formeln $(1 + r_t)(1 + r_{t-1})\dots(1 + r_0)$.

W	OMX minut 1/3		OMX minut 2/3		OMX minut 3/3		OMX minut 10.15-16.00					
							del 1/3		del 2/3		del 3/3	
m=	2	4	2	4	2	4	2	4	2	4	2	4
epsilon=												
0.5	9,34	11,71	10,30	10,39	12,85	12,65	9,05	11,09	8,88	8,51	10,96	10,40
0.8	11,27	12,73	12,28	12,37	14,63	14,01	10,55	11,33	10,29	9,60	13,62	12,37

Eftersom den asymptotiska fördelningen för detta värde är $N(0,1)$ framgår att sannolikheten att få ett lika stort eller större W_{me} , för andra mätvärden som verkligen är oberoende och likafördelade, bör vara mycket liten. Denna sannolikhet kallas p -värde. Det är tydligt att det finns en ganska stor säkerhetsmarginal för avvikelser från asymptotisk normalfördelning för BDS-testet. Enligt simuleringar gjorda av Brock, Hsieh och LeBaron (1993) är BDS-testets 1%-kvantiler vid 500 normalfördelade slumpantal för olika m och ε ²¹

$m =$	2	4
$\varepsilon =$		
0,5	-2,62; 3,00	-3,08; 3,64
1	-2,37; 2,61	-2,33; 2,87

Den ovan genomförda simuleringen antydde att BDS-testets uppförande med cauchyfordelade slumpantal åtminstone inte försämrades jämfört med normalfordelade slumpantal. Jag gör en (inte särskilt vågad) gissning om att residualerna från de finansiella tidsserierna har en fördelning någonstans mellan normalfördelning och cauchyfordelning. Eftersom jag har räknat med $\varepsilon = 0,8$ (som inte tabellerats av Brock et al) tvingas jag anta att $\varepsilon = 0,8$ inte har sämre (dvs till belopp större) kvantiler än det sämsta av $\varepsilon = 0,5$ och $\varepsilon = 1,0$. Med dessa antaganden är samtliga resultat i ovanstående resultattabell signifikant skilda från noll på 1%-nivån.

En annan intressant detalj är att alla W_{me} blev positiva (>0), men orsaken till detta är inte uppenbar. En rimlig gissning är att samma bakomliggande struktur är orsak till all påvisad avvikelse från IID.

Tabellerna nedan visar hur stort p -värdet är, under antagande att fördelningen istället är $N(0,1)$. Om p -värdet är mindre än 5% (= $5E-2$) är resultatet signifikant*.

W	Afgx månad		Afgx dag		Afgx dag utan helger	
	2	4	2	4	2	4
epsilon=						
0.5	9,E-08	<1E-15	9,E-05	1,E-08	7,E-05	8,E-10
0.8	1,E-07	<1E-15	3,E-04	6,E-09	2,E-04	4,E-09

²¹ Jfr normalfördelningskvantilen 1%: ± 2.58 .

W	OMX minut 10.15-16.00											
	OMX minut 1/3		OMX minut 2/3		OMX minut 3/3		del 1/3		del 2/3		del 3/3	
m=	2	4	2	4	2	4	2	4	2	4	2	4
epsilon=												
0.5	<1E-18	<1E-26	<1E-22	<1E-22	<1E-32	<1E-32	<1E-18	<1E-26	<1E-14	<1E-14	<1E-22	<1E-22
0.8	<1E-26	<1E-32	<1E-32	<1E-32	<1E-43	<1E-43	<1E-22	<1E-26	<1E-22	<1E-18	<1E-37	<1E-32

Vi ser att p -värdena är extremt små. Hypotesen om oberoende och likafördelning kan således förkastas med stor säkerhet, för alla serier, och dessa val av m och ε . Slutsatsen måste bli att den ovan beskrivna modellen (exponentiell tillväxt med oberoende, likafördelade avkastningar) är ofullständig. Ingen ledtråd ges emellertid beträffande vilken typ av modifiering av modellen som bör göras; det enda som är klart är att det finns information kvar i residualerna som istället borde finnas med i modellen.

Test för icke-linearitet i finansiella data

För att kunna göra uttalanden över eventuella icke-lineariteter måste först tillses att inget linjärt beroende finns i de data BDSstestet körs på. I Appendix IV visas därför utdrag ur en SPSS-körning, där det först utförts test (Box-Ljung) för autokorrelation.²² Något förvånande hittas signifikant autokorrelation i samtliga tidsserier. Den minsta ARMAmodell som tar bort all signifikant autokorrelation, och vars koefficienter för högsta ordningens termer är signifikant skilda från noll,²³ anpassas till tidsserierna. Residualerna uppvisar därefter ingen signifikant autokorrelation. De modeller som visade sig uppfylla detta var för serie 1a: ARMA(0,1), 2a: ARMA(4,4), 2b: ARMA(7,7), 3a: ARMA(6,6), 3b: ARMA(5,4). Det är också förvånande att så många termer kom med i modellerna. Det som är intressant för den fortsatta undersökningen i denna uppsats är emellertid endast att autokorrelation inte längre är ett problem. Residualerna från ARMA-modellerna bildar nu 5 nya tidsserier, motsvarande de 5 gamla 1a, 2a, 2b, 3a, 3b.

Dessa nya tidsserier körs in i BDSstestet. Om resultatet även nu visar att IID kan förkastas så kan slutsatsen bli att det finns icke-lineariteter i aktieindex.

²² SPSS är ett standardprogram för statistiska beräkningar som används främst inom de sociala vetenskaperna.

²³ Högsta ordningens koefficienter för ARMA(4,4) är de enda högsta ordningens koefficienter som inte blir signifikanta.

Resultat och slutsats

Nedanstående tabeller visar BDS-värdet $W_{m\epsilon}$ för några olika val av m och ϵ för de nya tidsserier som bildats då eventuell autokorrelation tagits bort.

	Afgx månad		Afgx dag		Afgx dag utan helger	
m=	2	4	2	4	2	4
epsilon=						
0.5	5,38	10,19	3,73	5,47	4,50	6,21
0.8	5,10	10,01	3,36	5,37	4,74	5,61

	OMX minut 1/3		OMX minut 2/3		OMX minut 3/3		OMX minut 10.15-16.00					
m=	2	4	2	4	2	4	del 1/3		del 2/3		del 3/3	
epsilon=							2	4	2	4	2	4
0.5	12,93	15,02	15,54	15,23	17,73	16,83	14,16	16,09	15,24	15,30	17,85	16,98
0.8	13,78	15,42	16,97	15,97	19,72	18,31	14,62	15,94	16,04	15,13	19,44	17,81

Man ser att alla testresultaten för OMXminut blivit större nu när eventuell autokorrelation tagits bort. Om detta beror på att avvikelser från IID blivit tydligare för BDS-testet, eller på att anpassningarna med ARMA-modellerna inte var så lyckade framgår emellertid inte. I övrigt avviker inte testresultaten speciellt mycket från de tidigare testerna (som utfördes när serierna hade signifikant autokorrelation). Samtliga tester i ovanstående tabell förkastar nollhypotesen: IID, med 99% approximativ signifikans, och detta kan alltså inte längre bero på autokorrelation.

Slutsatsen måste därför bli att det finns icke-linjärt beroende i aktieindexavkastning. Vad som orsakar detta är inte klart. Inte heller är det klart hur detta beroende ska komma med i den ursprungliga modellen. Som nämntes i avsnittet bakgrund är därför ARCH eller GARCH en rimlig första gissning.²⁴

²⁴ Det förtjänar att än en gång nämnas att även kaotiska system har icke-linjärt beroende i sig, men icke-linjärt beroende implicerar inte kaos.

Sammanfattning

BDS-testet för oberoende och likafördelning har beskrivits. Genom simulering har undersökts hur några av BDS-testets egenskaper förändras, när tidsserier med ovanligt tjocka svansar i fördelningsfunktionen testas. Detta har jämförts med BDS-testet för simulerade normalfördelade tidsserier. Resultatet blev att BDS-testets egenskaper inte försämras, tvärtom verkar de bli bättre. Alltså kan BDS-testet användas för att undersöka ekonomiska tidsserier.

BDS-testet har sedan använts för att undersöka oberoende och likafördelning i aktieindexavkastning. För samtliga undersökta tidsserier blev resultatet att hypotesen om *oberoende och likafördelning* kunde förkastas med stor säkerhet, för flera olika val av de 'godtyckliga' parametrarna. Slutsatsen blev att tidsserierna inte är IID.

Därefter togs eventuellt linjärt beroende bort från tidsserierna, och BDS-testet användes återigen. Nollhypotesen, som nu var *inget icke-linjärt beroende*, kunde förkastas med stor säkerhet för flera olika val av de 'godtyckliga' parametrarna. Slutsatsen blev att tidsserierna uppvisar icke-linearitet.

Källförteckning

- Abhyankar, A., Copeland, L. S., Wong, W. (1997), "Uncovering Nonlinear Structure In Real-Time Stock-Market Indexes: The S&P 500, the DAX, the Nikkei 225, and the FTSE-100", *Journal of Business & Economic Statistics*, 15, 1-14.
- Amilion, H., Byström, H. N. E. (1998), "The Search for Chaos and Nonlinearities in Swedish Stock Index Returns", Nationalekonomiska institutionen, Lunds universitet (1998, ännu ej publicerad?).
- Barnett, W. A., Gallant, A. R., Hinich, M. J., Jungeilges, J. A., Kaplan, D. T., Jensen, M. J. (1997) "A single-blind controlled competition among tests for nonlinearity and chaos", *Journal of Econometrics*, 82, 157-192.
- Brock, W. A. (1985), "Distinguishing Random and Deterministic Systems: Abridged Version", *Journal of Economic Theory*, 40, 168-195.
- Brock, W. A., Dechert, W. D., Scheinkman, J. A., LeBaron, B. (1996), "A Test for Independence based on the Correlation Dimension", *Econometric Reviews*, 15(3), 197-235.
- Day, R. H., "Complex Economic Dynamics: Obvious in History, Generic in Theory, Elusive in Data", *Nonlinear Dynamics, Chaos and Econometrics: An Introduction*, Hashem Pesaran, M., Potter, S. M. (1993), John Wiley & Sons Ltd, England.
- de Lima, P. J. F. (1996), "Nuisance parameter free properties of correlation integral based statistics", *Econometric Reviews*, 15(3), 237-259.
- Frank, M., Stengos, T. (1988), "The Stability of Canadian Macroeconomic Data as Measured by the Largest Lyapunov Exponent", *Economics Letters* 27(1), 11-14.
- Greene, W. H. (1997), *Econometric Analysis*, Prentice-Hall International, Inc.
- Kurths, J., Herzel, H. (1986), "An Attractor in a Solar Time Series", *Physica* 25D, 165-172.
- Lindgren, G., Rootzén, H. (1994), *Stationära stokastiska processer*, Institutionen för matematisk statistik, Lunds universitet.
- Liu, T., Granger, C. W. J., Heller, W. P., "Using the Correlation Exponent to Decide whether an Economic Series is Chaotic", *Nonlinear Dynamics, Chaos and Econometrics: An Introduction*, Hashem Pesaran, M., Potter, S. M. (1993), John Wiley & Sons Ltd, England.
- McCaffrey, Ellner, Gallant, Nychka (1992), "Estimating the Lyapunov Exponent of a Chaotic System with Nonparametric Regression", *Journal of the American Statistical Association*, 87(419), 682-695.
- Ohlén, G., Åberg, S. (1993), *Kaos inom naturvetenskap och teknik*, Institutionen för matematisk fysik, Lunds universitet.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., Flannery, B. P. (1992), *Numerical Recipes in C*, Cambridge University Press.

Rydén, T., Lindgren, G. (1996), *Markovprocesser*, Institutionen för matematisk statistik, Lunds universitet.

Stewart, I. (1990), *Does God Play Dice?*, Penguin Books.

Wolf, A., Swift, J. B., Swinney, H. L., Vastano, J. A. (1985), "Determining Lyapunov exponents from a time series", *Physica 16D*, 285-317.

Zeng, X., Pielke, R. A., Eykholt, R. (1992), "Extracting Lyapunov Exponents from Short Time Series of Low Precision", *Modern Physics Letters B*, Vol. 6, No. 2, 55-75.

Appendix I Matematiska definitioner

$\binom{a}{b}$ kallas *binomialkoefficient*, och anger antalet olika sätt att kombinera b st objekt ur a möjliga (utan hänsyn till ordningen, och utan att ha med samma objekt två eller fler gånger i en viss kombination). Binomialkoefficienterna beräknas $\binom{a}{b} = \frac{a!}{b!(a-b)!}$.

$\theta(a)$ benämns *stegfunktion*; den är 0 för $a \leq 0$, och 1 för $a > 0$. (Funktionen är egentligen inte definierad för $a=0$, men jag har använt ovanstående definition.)

Beteckningen $\|a\|$ står för *normen* av a , vilket är ett mått på storleken av a . Det finns oändligt många olika normer, men de vanligaste är ett-, två- och oändlighetsnormen ($\|a\|_1$, $\|a\|_2$, och $\|a\|_\infty$). För vektorer är det ganska enkelt att beräkna normer: 1-normen är den största kolumnsumman²⁵, oändlighetsnormen den största radsumman, och 2-normen är den geometriska längden av vektorn.²⁶ Max-normen är en sammanfattande benämning av bl a 1-normen för radvektorer och oändlighetsnormen för kolumnvektorer. Den beräknas genom att helt enkelt ta det största elementet.

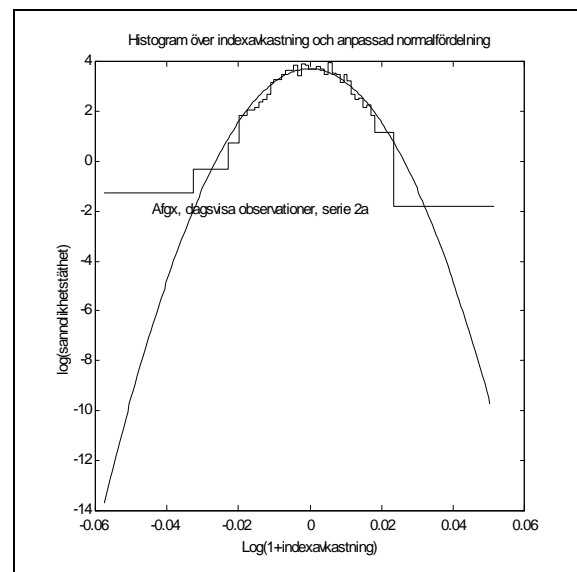
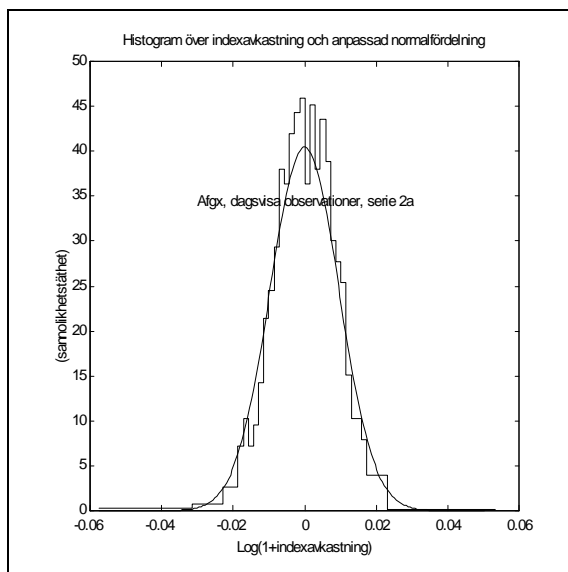
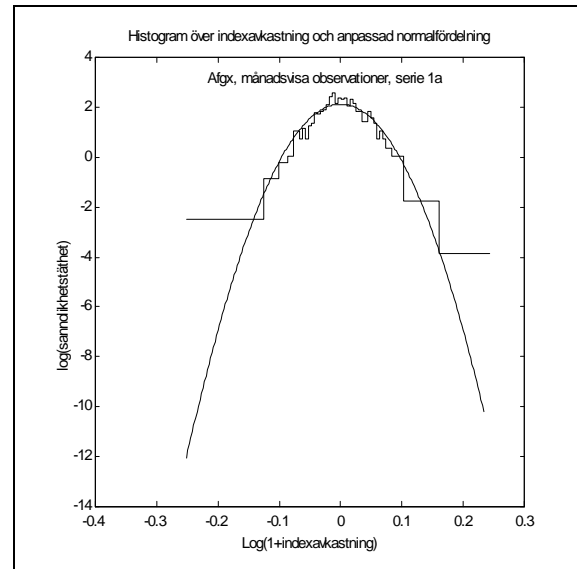
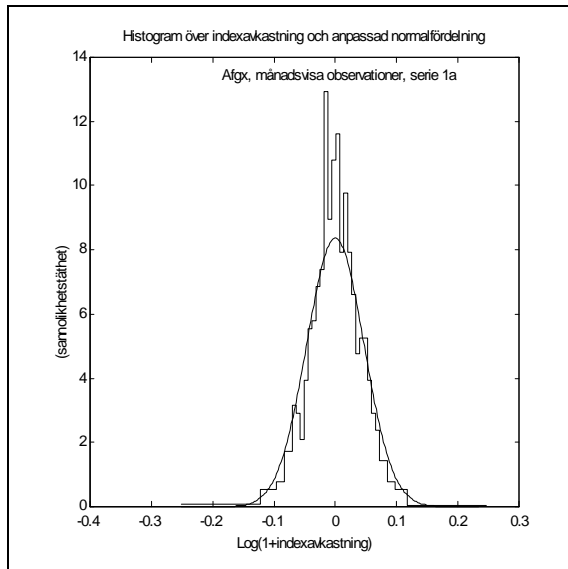
En *skalär* är ett tal, och kan tolkas som t ex en 1x1-matris eller en vektor med bara ett element.

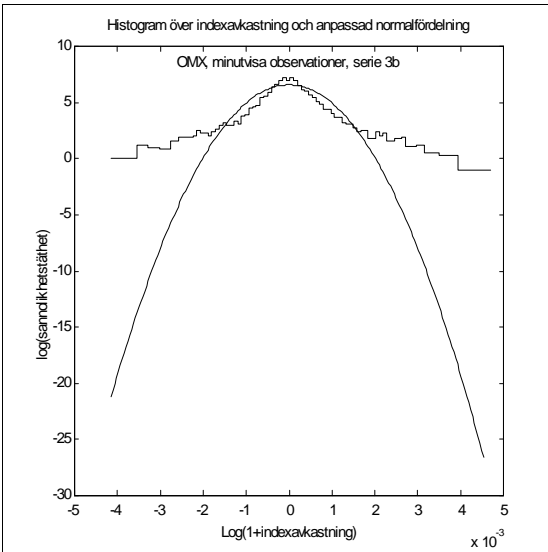
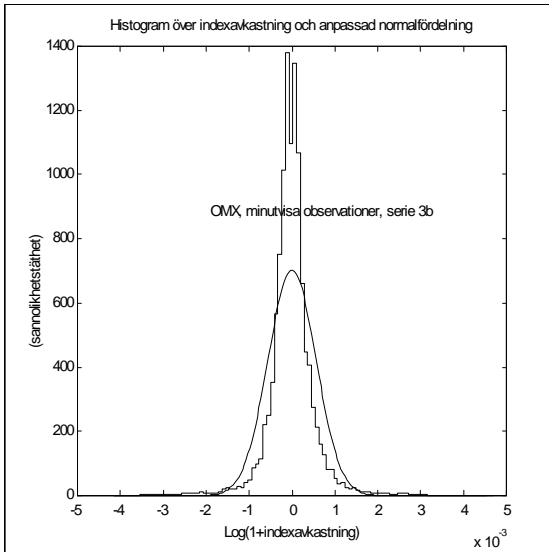
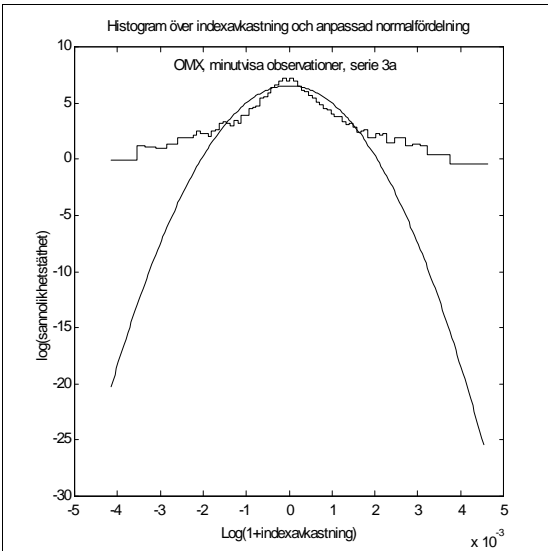
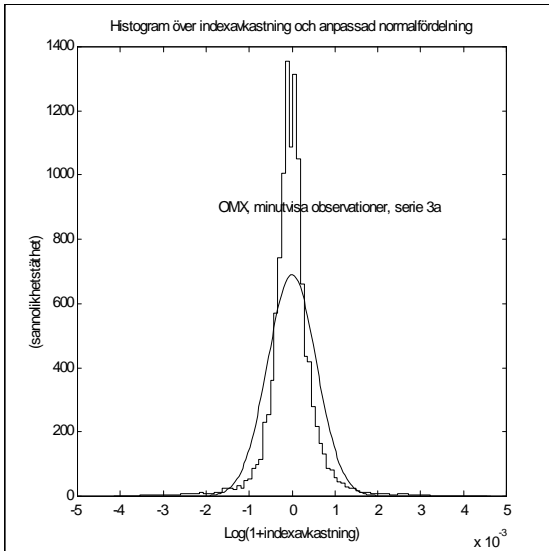
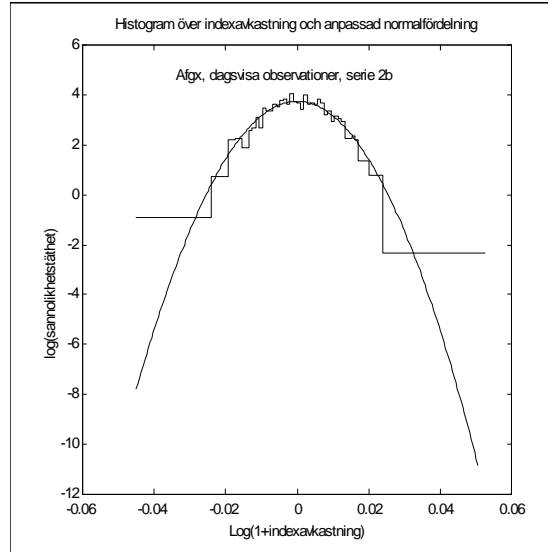
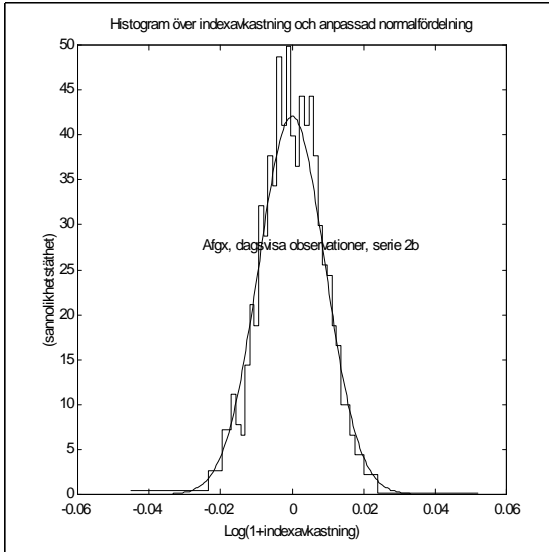
²⁵ Kolumnsumma = summan av absolutbeloppet av elementen i en kolumn. För en radvektor blir den största kolumnsumman helt enkelt det element som har störst absolutbelopp.

²⁶ Den geometriska längden av en vektor = roten ur(summan av (kvadraten av varje element))).

Appendix II Illustration av tjocka svansar

Följande histogram visar fördelningen för (logaritmerade) indexavkastningar. Minst 5 observationer finns i varje intervall (utom det sista). Lagg märke till att svansarna är tjockare än normalfördelningens, vilket syns tydligast på diagrammen till höger med logaritmerad sannolikhetstäthet.





Appendix III Matlabprogram för BDS-test

Bdsf.m:

```
% Detta program räknar ut ett bds-testvärde som är asymptotiskt  $N(0,1)$  om
mätdata är IID
% detta program fungerar för  $m \geq 2$ , liksom sats 2.2 i Brock et al (1996)
% function Wm=bdsf(m,epsilon,matdata);
% Mätdata bör ges som en kolumn.

function Wm=bdsf(m,epsilon,matdata);

% standardisera: (-> varians=1)
matdata=matdata/std(matdata);

n=length(matdata);

%bds-test

%  $C_m(m, \epsilon, \text{matdata}) = 1/n^2 \cdot \{\text{antal par } i, j \text{ sådana att: } ||X_{i,m} - X_{j,m}|| < \epsilon\}$ 
[Cm_m, C1, C, K]=Km(m,epsilon,matdata);

Tm=Cm_m-C1^m;

% beräkna variansen:
summan=0;
for j=1:m-1,
    summan=summan+C^(2*j)*(K^(m-j)-C^(2*m-2*j))-m*C^(2*m-2)*(K-C^2);
end;
Vm=sqrt(4*(m*(m-2)*C^(2*m-2)*(K-C^2)+K^m-C^(2*m)+2*summan));

% Wm är  $N(0,1)$  vid IID. Jämför abs(Wm) med t ex 1.96.
Wm=sqrt(n)*Tm/Vm;

% (Programmet anropar funktionen 'Km', se nedan)
% Så här räknar man ut hur lång tid detta program kommer att köra på en
dator med en AMD-K6-233 processor:
%  $2 \cdot (\text{antal mätpunkter}/400)^2 = \text{antal minuter}$ 
```

Km.m

```
% Programmet beräknar C och K som behövs för att beräkna bdstest-variansen.
% Dessutom beräknas  $C(m, \epsilon, \text{matdata})$  och  $C(1, \epsilon, \text{matdata})$ 
%  $K_m(m, \epsilon, \text{matdata}) \sim 1/n^3 \cdot \{\text{antal par } i, j, k \text{ sådana att: } ||X_{i,m} - X_{j,m}|| < \epsilon \text{ och } \dots\}$ 
function [yCm, yC1, y2, y1]=Km(m,epsilon,matdata);
n=length(matdata);

% skapa alla vektorer som kan behövas:
% en gång räcker ((för varje m))
%  $O(n)$ :
Xim=[];
for i=1:n-m+1,
    Xim(i,:)=matdata(i:i+m-1)'; % detta ger vektorerna som kolumner
end;
Xil=[];
for i=1:n-1+1,
    Xil(i,:)=matdata(i:i)'; % detta ger vektorerna som kolumner
end;

% beräkna alla normer och jämför med epsilon
```

```

% O(n^2):
integrandm=zeros(n-m+1);
for i=1:n-m+1,
    for j=1:n-m+1,
        if (norm(Xim(i,:)-Xim(j,:),inf)-epsilon)<0,
            integrandm(i,j)=1;
        end;
    end;
end;
integrandl=zeros(n-1+1);
for i=1:n-1+1,
    for j=1:n-1+1,
        if (norm(Xil(i,:)-Xil(j,:),inf)-epsilon)<0,
            integrandl(i,j)=1;
        end;
    end;
end;

% beräkna radsummorna:
% (är bra att ha senare)
% O(n^2):
radsummam=[];
radsummam=sum(integrandm,2); % radsumma (blir kolumnvektor)
radsummal=[];
radsummal=sum(integrandl,2);

% beräkna värdet av dubbelsumman:
% O(n):
antal1=sum(radsummal); % räkna antalet 1:or i matrisen

% beräkna värdet av trippelsumman:
% O(n^2):
antal=0;
for i=1:n-1+1,
    antal=antal+integrandl(i,:)*radsummal;
end;

% beräkna värdet av korrelationsintegralen C(m,epsilon,matdata) och
C(1,epsilon,matdata)
antalCm=0;
for i=1:n-m+1,
    antalCm=antalCm+sum(integrandm(i,i+1:n-m+1),2);
end; % dvs summera elementen ovanför (eller under) diagonalen
antalC1=0;
for i=1:n,
    antalC1=antalC1+sum(integrandl(i,i+1:n),2);
end;

y2=antal1/(n-1+1)^2;
y1=antal/(n-1+1)^3;

% här divideras inte med n^2 eftersom serien inte är 'tillräckligt stor'
yCm=antalCm/((m-n)*(m-n-1)/2);
yC1=antalC1/((1-n)*(1-n-1)/2);

```

Appendix IV SPSSkörning

Om slutsatsen ska kunna bli icke-linearitet eller inte efter ett BDS-test måste ev linjära komponenter tas bort innan BDS-testet körs, för att på så vis utesluta autokorrelation ur mothypotesen. Jag använder för detta SPSS, eftersom det ger snabb och förhållandevis smärtfri kontroll av autokorrelation och anpassning av ARIMA-modell. Utskriften nedan är starkt nedkortad.

Modell 1:

Test av autokorrelation visar, konstigt nog, att det finns linjär korrelation mellan residualerna.

Lösningen på denna autokorrelation blir därför att anpassa en lämplig ARIMA-modell och göra om autokorrelationstestet för residualerna från denna ARIMA-modell.

ACF

Auto- Stand.

Lag	Corr.	Err.	-1	-.75	-.5	-.25	0	.25	.5	.75	1	Box-Ljung	Prob.
1	.151	.041										13.433	.000

Nedanstående utskrift visar den ARIMA(0,0,1)-modell som visade sig signifikant för tidsserie 1.

Arima

DF Adj. Sum of Squares Residual Variance

Residuals	586	1.2482005	.00212994
-----------	-----	-----------	-----------

Variables in the Model:

	B	SEB	T-RATIO	APPROX. PROB.
MA1	-.15964196	.04078803	-3.9139418	.00010146

Nedanstående utskrift visar att residualerna från ARIMA(0,0,1)-modellen nu blivit fria från signifikant autokorrelation:

ACF

Auto- Stand.

Lag	Corr.	Err.	-1	-.75	-.5	-.25	0	.25	.5	.75	1	Box-Ljung	Prob.
1	-.003	.041										.006	.941
2	-.012	.041										.091	.956
3	.057	.041										2.044	.563
4	.003	.041										2.048	.727
5	-.007	.041										2.077	.838
6	.072	.041										5.134	.527
7	-.015	.041										5.264	.628
8	-.030	.041										5.785	.671
9	.041	.041										6.780	.660
10	.028	.041										7.262	.701
11	.049	.041										8.717	.648
12	.024	.041										9.074	.697
13	-.046	.041										10.367	.664
14	-.027	.041										10.817	.700
15	-.034	.041										11.523	.715
16	-.097	.041										17.260	.369

Tidsserie 2:

Test för autokorrelation:

ACF

Auto- Stand.

Lag	Corr.	Err.	-1	-.75	-.5	-.25	0	.25	.5	.75	1	Box-Ljung	Prob.
1	.050	.034										2.185	.139
2	-.004	.034										2.201	.333
3	-.006	.034										2.232	.526
4	-.112	.034										13.369	.010
5	-.023	.034										13.822	.017
6	.066	.034										17.656	.007
7	-.028	.034										18.365	.010
8	.021	.034										18.754	.016
9	.037	.034										19.941	.018
10	.018	.033										20.221	.027
11	.038	.033										21.529	.028
12	.001	.033										21.530	.043
13	.003	.033										21.539	.063
14	.018	.033										21.829	.082
15	-.021	.033										22.238	.102
16	-.047	.033										24.205	.085

Testet visar på en signifikant autokorrelation på tidsavståndet 4. Ingen signifikant ARIMAModell kan emellertid anpassas till dessa mätvärden, varför den detekterade autokorrelationen antagligen orsakats på slumpen. För att vara säker på att BDS-testet inte ger utslag för denna autokorrelation skapas ändå en ny serie av residualer till ARIMA(4,0,4):

Arima

DF Adj. Sum of Squares Residual Variance

Residuals	871	.08041741	.0009231
-----------	-----	-----------	----------

Variables in the Model:

	B	SEB	T-RATIO	APPROX. PROB.
AR1	.10577627	.23367274	.4526684	.65090026
AR2	-.39181951	.19310869	-2.0290103	.04276095
AR3	.50890930	.18847631	2.7001234	.00706583
AR4	-.51272908	.22539613	-2.2747909	.02316081
MA1	.05652390	.24714172	.2287105	.81914764
MA2	-.39396748	.20479678	-1.9236996	.05471788
MA3	.52345134	.20061097	2.6092857	.00922877
MA4	-.41150219	.23940169	-1.7188776	.08599199

Residualerna är nu åtminstone okorrelerade:

ACF

Auto- Stand.

Lag	Corr.	Err.	-1	-.75	-.5	-.25	0	.25	.5	.75	1	Box-Ljung	Prob.
1	.001	.034										.002	.966
2	-.007	.034										.040	.980
3	.027	.034										.679	.878
4	-.026	.034										1.267	.867
5	-.004	.034										1.283	.937
6	.057	.034										4.142	.657
7	-.005	.034										4.161	.761
8	-.007	.034										4.210	.838

9	.009	.034		.*	4.284	.892
10	.050	.033		.I*	6.472	.774
11	.024	.033		.*	6.967	.802
12	-.020	.033		.*	7.306	.837
13	.036	.033		.I*	8.488	.810
14	.010	.033		.*	8.578	.857
15	-.033	.033		*I.	9.576	.846
16	-.015	.033		.*	9.771	.878

Tidsserie 3:

Denna visar också signifikanta autokorrelationer:

ACF

Auto- Stand.

Lag	Corr.	Err.	-1	-.75	-.5	-.25	0	.25	.5	.75	1	Box-Ljung	Prob.
1	-.020	.037					.*					.277	.598
2	.082	.037					.I*					5.026	.081
3	-.132	.037					**I.					17.429	.001
4	-.020	.037					.*					17.722	.001
5	.043	.037					.I*					19.076	.002
6	-.089	.037					*I.					24.753	.000
7	.081	.037					.I*					29.441	.000
8	.001	.037					.*					29.442	.000
9	.027	.037					.I*					29.970	.000
10	-.009	.037					.*					30.029	.001
11	-.015	.037					.*					30.199	.001
12	-.016	.037					.*					30.374	.002
13	-.007	.037					.*					30.409	.004
14	-.030	.037					*I.					31.048	.005
15	.039	.037					.I*					32.150	.006
16	.037	.037					.I*					33.170	.007

ARIMA(7,0,7) anpassas: (vilket är något förvånande; sådana autokorrelationer bör inte kunna vara bestående över tiden)

Arima

DF Adj. Sum of Squares Residual Variance

Residuals	695	.05746982	.00008262
-----------	-----	-----------	-----------

Variables in the Model:

	B	SEB	T-RATIO	APPROX. PROB.
AR1	-.79257349	.23513019	-3.3707857	.00079100
AR2	-.75994012	.20946480	-3.6280087	.00030654
AR3	-.66637772	.28684785	-2.3231052	.02046123
AR4	-.48098052	.25422240	-1.8919675	.05891074
AR5	-.57678714	.24460446	-2.3580402	.01864777
AR6	-.69697848	.17453590	-3.9933244	.00007209
AR7	-.62239513	.15130804	-4.1134306	.00004365
MA1	-.79866199	.23366439	-3.4179876	.00066765
MA2	-.83934133	.22589954	-3.7155514	.00021905
MA3	-.58409573	.31684164	-1.8434942	.06568245
MA4	-.44253513	.27779121	-1.5930494	.11160377
MA5	-.56967663	.27311336	-2.0858614	.03735531
MA6	-.58552044	.20006225	-2.9266913	.00353773
MA7	-.65314741	.15249772	-4.2829978	.00002104

Residualerna blir åtminstone fria från autokorrelation:

ACF

Auto- Stand.

Lag	Corr.	Err.	-1	-.75	-.5	-.25	0	.25	.5	.75	1	Box-Ljung	Prob.
1	.000	.037					.*					.000	.993
2	.022	.037					.*					.348	.840
3	-.013	.037					.*					.464	.927
4	-.018	.037					.*					.703	.951
5	-.001	.037					.*					.704	.983
6	.014	.037					.*					.855	.991
7	-.002	.037					.*					.857	.997
8	-.021	.037					.*					1.185	.997
9	.012	.037					.*					1.286	.998
10	-.022	.037					.*					1.620	.999
11	.028	.037					.I*					2.192	.998
12	-.011	.037					.*					2.281	.999
13	-.024	.037					.*					2.681	.999
14	.027	.037					.I*					3.199	.999
15	.025	.037					.*					3.636	.999
16	.000	.037					.*					3.636	.999

Tidsserie 4:

Autokorrelationen är mycket tidlig för minutdata (korrelationen avtar dock snabbt):

ACF

Auto- Stand.

Lag	Corr.	Err.	-1	-.75	-.5	-.25	0	.25	.5	.75	1	Box-Ljung	Prob.
1	-.267	.010					*****I					760.687	.000
2	.011	.010					*					761.989	.000
3	.035	.010					I*					774.779	.000
4	.042	.010					I*					793.659	.000
5	.020	.010					*					797.745	.000
6	.002	.010					*					797.776	.000

Anpassning av ARIMAModell: ARIMA(6,0,6) ger signifikanta högstakoefficienter.

Arima

DF Adj. Sum of Squares Residual Variance

Residuals	10642	.00312692	.00000029
-----------	-------	-----------	-----------

Variables in the Model:

	B	SEB	T-RATIO	APPROX. PROB.
AR1	.8723589	.18800006	4.6402054	.00000352
AR2	.5694927	.17690944	3.2191200	.00128970
AR3	-.4542639	.20734346	-2.1908766	.02848233
AR4	-.7081227	.17039047	-4.1558820	.00003266
AR5	.6398616	.14230596	4.4963794	.00000699
AR6	-.1317831	.03506465	-3.7582882	.00017199
MA1	1.1630092	.18784224	6.1914144	.00000000
MA2	.2909944	.20715080	1.4047469	.16012575

MA3 -.6440285 .21414056 -3.0075038 .00264018
 MA4 -.5743911 .17904382 -3.2081032 .00134009
 MA5 .8828716 .17723815 4.9812728 .00000064
 MA6 -.3151211 .06110241 -5.1572607 .00000026

Residualerna undersöks nedan. Ingen korrelation uppvisas:

ACF

Auto- Stand.

Lag	Corr.	Err.	-1	-.75	-.5	-.25	0	.25	.5	.75	1	Box-Ljung	Prob.
1	.000	.010					*					.002	.962
2	-.002	.010					*					.056	.972
3	.003	.010					*					.160	.984
4	-.001	.010					*					.174	.996
5	-.003	.010					*					.246	.999
6	-.002	.010					*					.308	.999
7	.003	.010					*					.396	1.000
8	.001	.010					*					.406	1.000
9	.000	.010					*					.407	1.000
10	-.001	.010					*					.428	1.000
11	-.008	.010					*					1.045	1.000
12	-.005	.010					*					1.284	1.000
13	.002	.010					*					1.314	1.000
14	.015	.010					*					3.833	.996
15	.002	.010					*					3.887	.998
16	-.010	.010					*					4.894	.996

Tidsserie 5:

Test för autokorrelation:

ACF

Auto- Stand.

Lag	Corr.	Err.	-1	-.75	-.5	-.25	0	.25	.5	.75	1	Box-Ljung	Prob.
1	-.309	.010					*****I					967.374	.000
2	-.007	.010					*					967.906	.000
3	.012	.010					*					969.481	.000
4	.025	.010					I*					975.980	.000
5	.003	.010					*					976.061	.000
6	-.009	.010					*					976.876	.000
7	.016	.010					*					979.495	.000

En ARIMA(5,0,4)modell visar sig vara signifikant:

Arima

DF Adj. Sum of Squares Residual Variance

Residuals 10152 .00279658 .00000028

Variables in the Model:

	B	SEB	T-RATIO	APPROX. PROB.
AR1	-.72664021	.10516533	-6.9095035	.00000000
AR2	.31519603	.10193274	3.0921960	.00199220
AR3	.77818692	.10186091	7.6397013	.00000000
AR4	.13897475	.09550725	1.4551225	.14566627
AR5	.06189057	.02813141	2.2000519	.02782566
MA1	-.37956933	.10519035	-3.6084044	.00030957
MA2	.56835544	.09328481	6.0926902	.00000000
MA3	.64613365	.09665535	6.6849235	.00000000
MA4	-.17321661	.08583892	-2.0179263	.04362519

Test av residualerna från denna ARIMA-modell visar att de nu är okorrelerade:

ACF

Auto- Stand.

Lag	Corr.	Err.	-1	-.75	-.5	-.25	0	.25	.5	.75	1	Box-Ljung	Prob.
1	.000	.010					*					.000	.995
2	.000	.010					*					.001	1.000
3	.000	.010					*					.001	1.000
4	.004	.010					*					.135	.998
5	.000	.010					*					.137	1.000
6	-.007	.010					*					.587	.997