

Icke-linjära tidsseriemodeller

1 Egenskaper hos finansiella tidsserier

Innan vi introducerar icke-linjära modeller och ser på (den något diffusa) avgränsningen från linjära modeller, ser vi i avsnitt 1.1-1.3 på några vanligt förekommande egenskaper hos finansiella tidsserier. Dessa egenskaper kan inte förklaras med linjära modeller och visar därför på nödvändigheten av icke-linjära tidsseriemodeller.

1.1 Stora förändringar är vanligare än förväntat

Låt y_t vara diff. $\log^1$ av P_t :

$$y_t = 100(\ln P_t - \ln P_{t-1})$$

där P_t är priset på t.ex. en aktie eller en valuta vid tidpunkt t . Ofta antas att serien y_t är oberoende och identiskt normalfördelad, dvs.

$$y_t \sim \text{o.i.f.} N(\mu, \sigma^2)$$

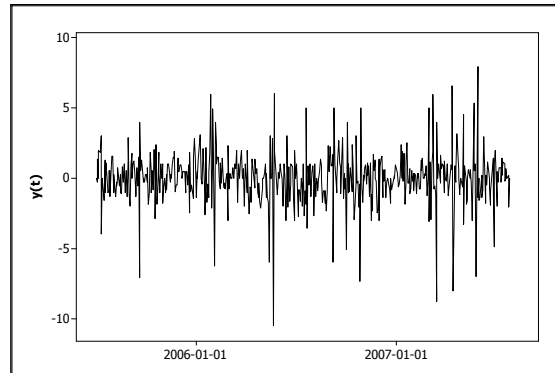
Under detta antagande har y_t skevhet 0 och toppighet 3.

En genomgång av åtta aktiemarknader (Amsterdam, Frankfurt, Hong Kong, London, New York, Paris, Singapore, Tokyo) dag för dag under perioden 2 januari 1980 till 31 december 1997, visar genomgående en negativ skevhet och en toppighet större än 3. Dvs. stora negativa förändringar är vanligare än stora positiva, och dessa förändringar är vanligare än förväntat under normalitetsantagandet. Figur 1.1, 1.2 och tabell 1.1 nedan visar att detta även gäller för Stockholmsbörsen, där $y_t = 100(\ln P_t - \ln P_{t-1})$ och där P_t

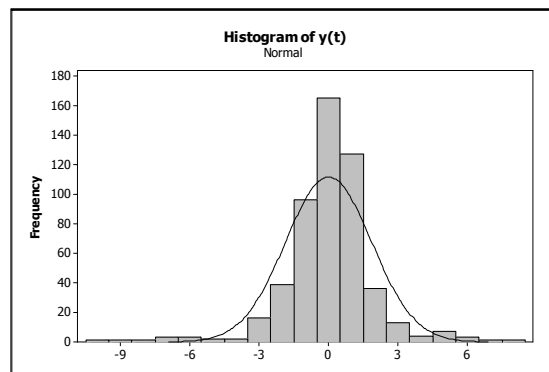
¹För stora tal är diff. $\log \times 100 \approx$ procentuell förändring.

är OMX-S slutkurs dag t perioden 2005-07-01 till 2007-07-25 uttryckt i US dollar. I tabell 1.1 redovisas dessutom de observerade måttens standardfel samt de förväntade värdena givet dragning från $N(0, 1)$.

Figur 1.1: Diff log för Stockholmsbörsen 2005-07-01 till 2007-07-25



Figur 1.2: Histogram över diff log för Stockholmsbörsen 2005-07-01 till 2007-07-25



Tabell 1.1: Beskrivande mått för diff log för Stockholmsbörsen 2005-07-01 till 2007-07-25

	OMX-S	S.E.	Förväntat
Medel	0.032	0.082	0
Median	0	0.102	0
Min	-10.492		
Max	7.950		
Varians	3.479		1
Skevhet	-0.633	0.107	0
Toppighet	5.749	0.215	3

Stockholmsbörsen har under de senaste två åren i genomsnitt gått upp med 0.032 procent per dag. Skevheten är signifikant nollskild och toppigheten är signifikant skild från tre. Detta kan också ses i figur 1.1 och 1.2, där stora negativa värden är vanligare än stora positiva, och där värden nära noll är vanligare än förväntat under antagandet om normalfördelning.

Toppigheten är fortfarande större än 3 om man studerar åtta stora valutors växelkurs mot US dollarn (Australiska dollar, Brittiska pund, Kanadensiska dollar, Holländska gulden, Franska franc, Tyska D-mark, Japanska Yen, Schweiziska franc), men dessa serier är i det närmaste symmetriska (skevheten är nära noll).

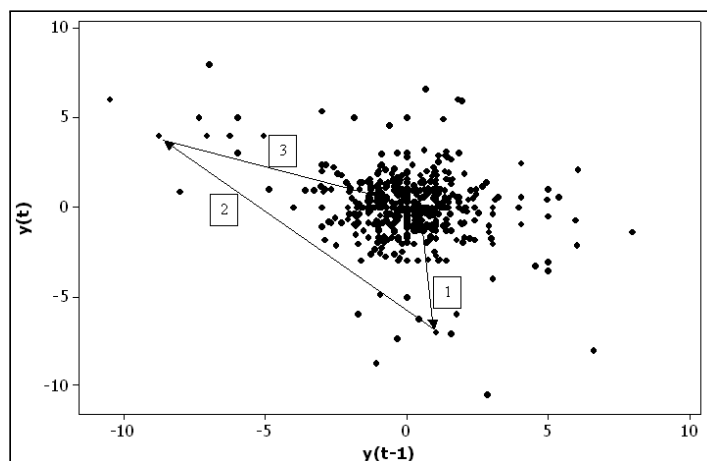
1.2 Heteroskedasticitet

I figur 1.1 ser det ut som att stora förändringar kommer i kluster, dvs. avkastningen verkar vara heteroskedastisk. Detta styrks också av ARCH-LM testet för heteroskedasticitet (p-värde²: 0.0082). En stor negativ avkastning tenderar alltså att efterföljas av en stor positiv och vice versa. Detta faktum gäller även för de åtta aktiekurserna och valutorna som studerades i avsnitt 1.1.

Ett intressant och ofta förekommande fenomen i framförallt aktiemarknader är den sk. "tre-cykeln". Om vi plottar y_t mot y_{t-1} för OMX-S fås:

²För nollhypotesen om homoskedasticitet.

Figur 1.3: Plot över avkastningen på Stockholmsbörsen med illustration av "tre-cykeln"



Tre-cykeln innebär att då väl en observation lämnar "huvudmolnet" runt ($y_t = 0, y_{t-1} = 0$) efterföljs denna observation av ytterligare en observation utanför molnet innan återgång till huvudmolnet.

1.3 Negativ avkastning efterföljs ofta av en period av hög volatilitet

I figur 1.3 finns ingen tidsskala, men faktum är att avvikelser från huvudmolnet nästan uteslutande sker i syddlig riktning. Tre-cykel fenomenet gör att det ofta tar tre på varandra följande tidpunkter innan återgång till huvudmolnet, dvs. en volatil period inleds ofta med en stor negativ avkastning. Ytterligare ett tecken på detta är att $Corr(y_t^2, y_{t-1})$ uteslutande är negativ för de åtta stora aktiemarknaderna och för Stockholm OMX-S (-0.063). Detsamma gäller dock inte för valutamarknaderna.

Hur tidsserier med dessa egenskaper ska modelleras står beskrivet i följande text. Som brevidläsningslitteratur rekommenderas t.ex. Granger och Teräsvirta (1993) samt Franses och van Dijk (2000).

2 Introduktion

Avsnitt 1.1-1.3 tyder på att finansiella data är asymmetriska. Om denna asymmetri överhuvudtaget kan förklaras måste svaret finnas bland de icke-linjära tidsseriemodellerna, vilket är ett mycket aktuellt forskningsområde. För att förklara vad som menas med en icke-linjär process definieras först den linjära processen

$$y_t = \alpha + \sum_{i=0}^{\infty} \psi_i a_{t-i} \quad (2.1)$$

där α är en konstant, ψ_i är reella tal med $\psi_0 = 1$ och $\{a_t\}$ är en sekvens av oberoende lika fördelade stokastiska variabler med en definierad fördelningsfunktion. Vidare antas a_t vara kontinuerlig och $E(a_t) = 0$. Ofta antas dessutom att $\text{Var}(a_t) = \sigma_a^2$, eller ännu starkare - att a_t är gaussisk. Om $\sigma_a^2 \sum_{i=1}^{\infty} \psi_i^2 < \infty$, är y_t svagt stationär. (Stationära) ARMA-processer är linjära eftersom de har en MA-representation.

En linjär *prognosmetod* är en metod där prognosen för tidpunkt $N + h$ kan uttryckas som en linjär funktion av de observerade värdena upp till och med N . Exempel på linjära prognosmetoder är enkel exponentiell utjämning, Holts metod och additiva (men inte multiplikativa) Holt-Winters metod. Dessa linjära metoder för exponentiell utjämning har också tydliga kopplingar till ARIMA-modeller. Den enkla exponentiella utjämningen är ekvivalent med en ARIMA(0, 1, 1)-modell där $\theta = 1 - \alpha$, och där α är utjämningskonstanten. Holts metod är ekvivalent med en ARIMA(0, 2, 2)-modell där $\theta_1 = 2 - \alpha - \alpha\beta$ och $\theta_2 = \alpha - 1$, där α och β är de två utjämningskonstanterna. Vidare är den additiva Holt-Winters metoden ekvivalent med en ARIMA(0, 1, $s + 1$)(0, 1, 0) $_s$ -modell. Den multiplikativa Holt-Winters är icke-linjär och har därför ingen ekvivalent ARIMA-modell. Däremot är multiplikativa Holt-Winters metoden sk. log-linjär, dvs. logaritmen av serien med multiplikativ säsongvariation kan beskrivas med linjära modeller.

En process som inte uppfyller ekvation (2.1) sägs vara icke-linjär. Låt y_t istället uttrycks i termer av dess betingade moment, där F_{t-1} är beteckningen för all tillgänglig information vid tidpunkt $t - 1$, dvs, F_{t-1} är en samling av linjärkombinationer av elementen i $\{y_{t-1}, y_{t-2}, \dots\}$ och $\{a_{t-1}, a_{t-2}, \dots\}$. Det betingade medelvärde och variansen av y_t givet F_{t-1} är

$$\mu_t = E(y_t | F_{t-1}) \equiv g(F_{t-1}) \quad (2.2)$$

och

$$\sigma_t^2 = \text{Var}(y_t | F_{t-1}) \equiv h(F_{t-1}) \quad (2.3)$$

där g och h är väldefinierade funktioner med $h > 0$. Vi begränsar alltså modellen till

$$y_t = g(F_{t-1}) + \sqrt{h(F_{t-1})}\epsilon_t$$

där $\epsilon_t = a_t/\sigma_t$. För den linjära serien y_t i ekvation (2.1) är $g(\cdot)$ en linjär funktion av element i F_{t-1} , och $h(\cdot) = \sigma_a^2$. I utvidgningen till icke-linjära modeller tilläts icke-linjära funktioner i ekvation (2.2) och (2.3). Om $g(\cdot)$ är icke-linjär, sägs y_t vara *icke-linjär i medelvärde*. Om $h(\cdot)$ är tidsberoende, sägs y_t vara *icke-linjär i varians*. Här kommer denna uppdelning att följas. Efter ett inledande kapitel om några icke-linjära utvecklingar av ARMA-processer behandlas i kapitel 4 modeller som är icke-linjära i medelvärde. Kapitel 5 behandlar modeller för tidsserier som är icke-linjära i varians. Denna uppdelning är fullgod, men (som alltid) finns undantag. Här är undantaget GARCH-M modeller där μ_t beror på σ_t^2 . Jag har valt att behandla GARCH-M i kapitel 5 bland alla andra ARCH-släktingar. I det avslutande kapitlet 6 behandlas de flexibla artificiella neurala nätverks (ANN)-modellerna. Kapitel 7 sammanfattar.

3 Icke-linjära former av ARMA-processer

Den enklaste icke-linjära utvecklingen av en ARMA-process är den sk. NLAR (NonLinear AutoRegressive)-modellen

$$y_t = f(y_{t-1}) + \varepsilon_t \quad (3.1)$$

vilket är en första ordningen icke-linjära autoregressiv modell, NLAR(1). Notera att feltermerna adderas till den icke-linjära funktionen (till skillnad från t.ex. de bilinjära modellerna, se nedan). Låt säga att $f(y_{t-1}) = \psi_1 y_{t-1}^2$, då skrivs NLAR(1)-modellen

$$y_t = \psi_1 y_{t-1}^2 + \varepsilon_t \quad (3.2)$$

Generellt skrivs den p :te ordningen icke-linjära autoregressiva modellen

$$y_t = f(y_{t-1}, y_{t-2}, \dots, y_{t-p}) + \varepsilon_t \quad (3.3)$$

och betecknas NLAR(p). Svårigheten att skatta (3.3) ligger i att den funktionella formen av $f(\cdot)$ är okänd. Ett sätt att komma vidare är att Taylor-approximera den okända formen. T.ex. blir för NLAR(2)-modellen, $y_t = f(y_{t-1}, y_{t-2}) + \varepsilon_t$

$$\begin{aligned} y_t = & \alpha + \psi_1 y_{t-1} + \psi_2 y_{t-2} + \psi_{12} y_{t-1} y_{t-2} + \psi_{11} y_{t-1}^2 + \psi_{22} y_{t-2}^2 + \\ & \psi_{112} y_{t-1}^2 y_{t-2} + \psi_{122} y_{t-1} y_{t-2}^2 + \psi_{111} y_{t-1}^3 + \psi_{222} y_{t-2}^3 + \varepsilon_t \end{aligned}$$

För den generella NLAR(p)-modellen behövs mer kompakta beteckningar. Ett sätt att skriva dess Taylor-approximation är

$$y_t = \alpha + \sum_{i=1}^p \psi_i y_{t-i} + \sum_{i=1}^p \sum_{j=1}^p \sum_{k=1}^r \sum_{l=1}^s \psi_{ijkl} y_{t-i}^k y_{t-j}^l + \varepsilon_t \quad (3.4)$$

där p är processens ordning, r och s är heltal ≥ 1 . Ekvation (3.4) är en sk. GAR (Generalized AutoRegressive)-process av ordning p . GAR-processer introducerar olika potenser av y_{t-i} och y_{t-j} samt dess kryssprodukter. GAR-processer är därför en naturlig utveckling av AR-processer. Det är kvadrupelsumman i (3.4) som gör processen icke-linjär, och som utgör skillnaden mellan GAR(p) och den linjära AR(p)-processen.

En annan naturlig icke-linjär utveckling av ARMA-processer är att bilda kryssprodukter av y_{t-i} och ε_{t-j} . Den enklaste formen av dessa sk. Bilinjära (BL)-processer är

$$y_t = \alpha + \psi_1 y_{t-1} + \omega_1 \varepsilon_{t-1} + \gamma_1 y_{t-1} \varepsilon_{t-1} + \varepsilon_t \quad (3.5)$$

eller

$$y_t = \alpha + (\psi_1 + \gamma_1 \varepsilon_{t-1}) y_{t-1} + \varepsilon_t^* \quad (3.6)$$

vilken betecknas BL(1, 1, 1, 1), där $\varepsilon_t^* = \varepsilon_t + \omega_1 \varepsilon_{t-1}$. Ekvation (3.6) ser ut som en AR(1)-process med AR-parameter, $\psi_1 + \gamma_1 \varepsilon_{t-1}$. BL(1, 1, 1, 1) kan alltså ses på som en AR(1)-modell, med stokastiskt varierande AR-parameter. Den stokastiska autoregressiva koefficienten har förstås väntevärdet ψ_1 . Om γ_1 är positiv, ökar koefficienten med ε_{t-1} , dvs. positiva

störningar ($\varepsilon_{t-1} > 0$) är mer persistenta än negativa störningar. I teorin bör bilinjära modeller vara särskilt lämpade för beskriva aktiekurser där just positiva störningar anses vara mer persistenta än negativa störningar. Som nämntes i kap. 1 anses dessutom stora negativa förändringar vara vanligare än stora positiva, något som också kan genereras med bilinjära modeller (se simuleringstudien nedan).

Den generella formen av den bilinjära processen $BL(p, q, r, s)$ är

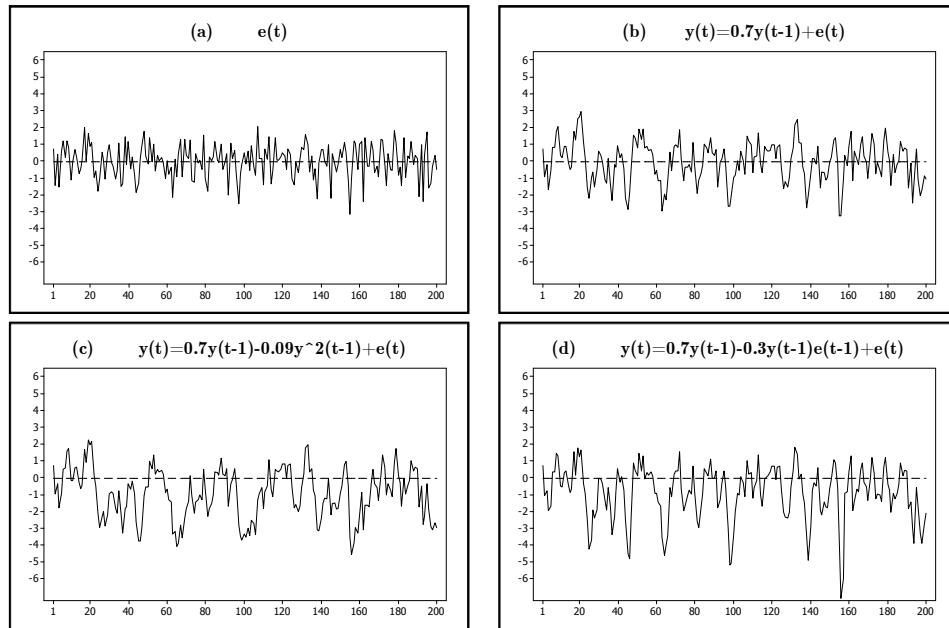
$$y_t = \alpha + \sum_{i=1}^p \psi_i y_{t-i} + \sum_{j=1}^p \omega_j \varepsilon_{t-j} + \sum_{i=1}^r \sum_{j=1}^s \gamma_{ij} y_{t-i} \varepsilon_{t-j} + \varepsilon_t \quad (3.7)$$

Notera att då $\gamma_{ij} = 0$, reduceras $BL(p, q, r, s)$ till den linjära $ARMA(p, q)$ -processen.

Simuleringsstudie

I syfte att bättre förstå dessa icke-linjära utvecklingar av $ARMA$ -processer visas i figur 3.1 resultatet av fyra simulerade processer.

Figur 3.1(a-d) Simulering av vitt brus, $AR(1)$, $GAR(1)$ och $BL(1, 0, 1, 1)$



I figur 3.1(a) ($e(t)$) visas 200 simulerade observationer dragna från $N(0, 1)$, på vilka de tre andra simulerade processerna bygger. I figur 3.1(b) visas en simulerad $AR(1)$ -process med $\phi = 0.7$, där y_1 satts lika med e_1 .

I figur 3.1(c) ses resultatet från den simulerade GAR(1)-processen

$$y_t = 0.7y_{t-1} - 0.09y_{t-1}^2 + \varepsilon_t$$

eller

$$y_t = (0.7 - 0.09y_{t-1})y_{t-1} + \varepsilon_t$$

Den här GAR(1)-processen uppför sig som en AR(1)-process med varierande koefficient, då y_{t-1} ökar så minskar AR-koefficienten. T.ex. för värdena $y_{t-1} = -2, 0$ och 2 är den autoregressiva persistensen $0.88, 0.7$ respektive 0.52 . Skillnaden i persistens ses tydligt i figur 3.1(c); de negativa värdena av den simulerade GAR-processen är betydligt mer persistenta än de positiva. Denna egenskap gör att GAR-modeller inte är lämpade för att beskriva finansiella tidsserier, där istället positiva störningar anses vara mer persistenta än negativa störningar. I en jämförelse mellan GAR-processen med AR(1)-processen (som har persistens 0.7 för alla värden) så är det tydligt att GAR-processen har en betydligt långsammare återgång mot processens medelvärde (t.ex. runt observation 100 och 160). Detta fenomen gör förstås också att negativa värden är vanligare än positiva, vilket inte heller är önskvärt vid modellering av finansiella tidsserier.

I figur 3.1(d) ser vi den sekvens som genererats från den bilinjära processen

$$y_t = 0.7y_{t-1} - 0.3y_{t-1}\varepsilon_{t-1} + \varepsilon_t$$

eller

$$y_t = (0.7 - 0.3\varepsilon_{t-1})y_{t-1} + \varepsilon_t$$

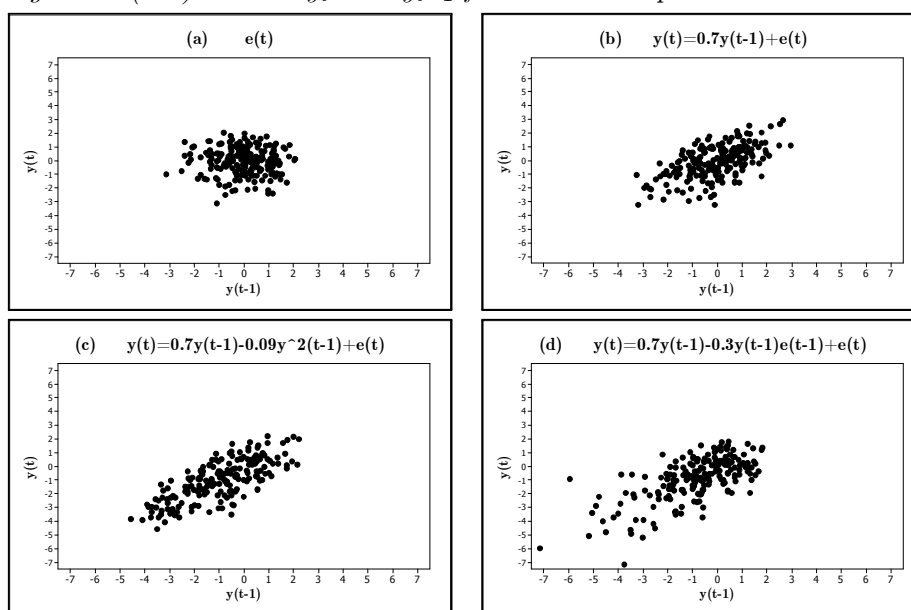
I den bilinjära modellen beror persistensen på värdet ε_{t-1} , ju större värde desto mindre persistens. För värden $\varepsilon_{t-1} < -1$, är processen explosiv. I tabell 3.1 ses beskrivande statistik för de fyra simulerade processerna.

Tabell 3.1 Beskrivande statistik för simulerade processer

	Vitt brus	AR(1)	GAR(1)	BL(1, 0, 1, 1)
Medel	-0.02	-0.06	-0.96	-0.88
Median	0.06	0.02	-0.88	-0.66
Min	-3.13	-3.25	-4.55	-7.17
Max	2.06	2.96	2.22	1.82
Varians	0.99	1.49	2.19	2.55
Skevhet	-0.38	-0.27	-0.17	-0.98
Toppighet	2.86	2.84	2.34	4.14
ARCH-LM (p-värde)	0.47	0.12	0.00	0.00
Jarque-Bera (p-värde)	0.18	0.26	0.10	0.00

Den simulerade GAR serien, och kanske framförallt den simulerade bilinjära serien verkar lovande med tanke på att vi från kap. 1 och ekonomisk teori vet att finansiella tidsserier har negativ skevhet, toppighet större än tre och att stora förändringar är vanligare än förväntat. Vi ser också att de simulerade GAR- och bilinjära processerna är tydligt heteroskedastiska. I figur 3.2 visas första ordningens beroende för de fyra simulerade processerna.

Figur 3.2 (a-d) Plottar y_t mot y_{t-1} för simulerade processer



Korrelationen mellan y_t och y_{t-1} (r_1) är för den simulerade vitt brus-processen i 3.2(a) lika med -0.03 . För de tre andra simulerade processerna (AR(1),

GAR(1) och BL(1, 0, 1, 1)) är korrelationen 0.59, 0.76 resp. 0.68. Att r_1 är störst för GAR processen är inte förvånande eftersom för denna process är persistensen störst (större än 0.7) för negativa värden, och de flesta värdena är negativa. Förväntad persistens för vitt brus-processen är förstås 0, och för AR -och BL processen 0.7.

En annan möjlighet uppstår om vi antar att f i ekvation (4.3) är bitvis linjär. Om dessa viktiga icke-linjära processer handlar kapitel 4.

4 Icke-linjäritet i medelvärde

Låt säga att f i NLAR(1)-processen $y_t = f(y_{t-1}) + \varepsilon_t$, är delvis linjär. Parametrarna i f skulle då t.ex. kunna bero på huruvida y_{t-1} är större eller mindre än ett visst kritiskt värde. Detta kritiska värde brukar kallas *tröskelvärde* (threshold), därav kallas dessa delvis linjära processer för *tröskelprocesser*. Dessa tröskelprocesser kallas också för *regimskiftesmodeller* där tidsseriens medelvärde, varians och autokorrelation kan vara olika i olika regimer. T.ex. visade Le Baron (1992) att autokorrelationerna i aktieavkastning är beroende av volatiliteten i avkastningen. Autokorrelationerna var stora i perioder av låg volatilitet och vice versa. Perioderna av låg och hög volatilitet kan här tolkas som olika regimer, eller uttryckt annorlunda, graden av volatilitet är den regimbestämmande processen. Ett annat klassiskt exempel på en regimskiftesprocess är hur andelen arbetslösa påverkas av konjunkturen. I en konjunkturedgång brukar andelen arbetslösa öka kraftigt till en början, för att sedan långsamt gå ner. I en konjunkturuppgång däremot sjunker inte andelen arbetslösa lika kraftigt. Dynamiken för andelen arbetslösa förändras alltså beroende på om ekonomin är på väg upp eller ned. Ekonomins status skulle här kunna bilda grund för två eller fler regimer.

Lite grovt kan dessa regimskiftesmodeller delas in två klasser. I den första klassen antas regimerna bestämmas av en observerbar variabel. Dvs. vi vet med säkerhet vilka regimskiften som har inträffat. I den andra klassen av regimskiftesmodeller antas regimerna bestämmas av en icke observerbar stokastisk process. Vi kan aldrig vara säkra på om ett särskilt regimskifte inträffat i en viss tidpunkt. Däremot går det att beräkna sannolikheter för de olika regimerna.

De två klasserna av regimskiftesmodeller behandlas separat i delkapitel 4.1 och 4.2. Olika test för icke-linjäritet redovisas i avsnitt 4.3. Hur dessa modeller estimeras, diagnosticeras och prognosticeras behandlas i avsnitt 4.4-4.6.

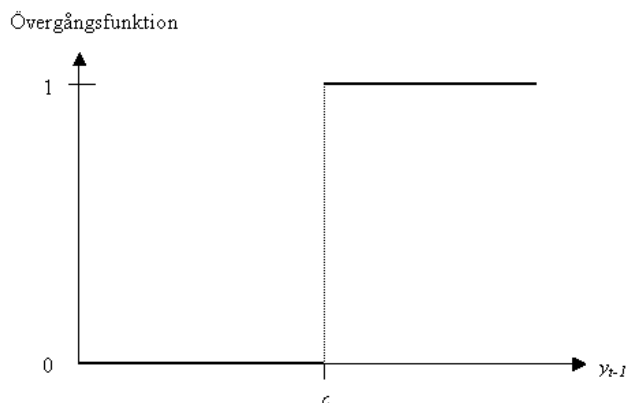
4.1 Modeller där regimerna bestäms av observerbar variabel

Den kanske enklaste modellen i denna klass är den sk. TAR (Threshold Autoregressive)-modellen som introducerades av Tong (1978). I denna modell antas regimerna bestämmas av den observerbara tröskelvariabeln q_t i relation till tröskelvärde c . Ett specialfall uppstår då $q_t = y_{t-d}$, där d är ett heltal större än 0. De modeller i vilka regimerna bestäms av tidsserien själv brukar kallas SETAR (Self-Exciting TAR)-modeller. T.ex. då $d = 1$, och en AR(1)-modell antas i två regimer fås

$$y_t = \begin{cases} \phi_{0,1} + \phi_{1,1}y_{t-1} + \varepsilon_t & \text{om } y_{t-1} \leq c \\ \phi_{0,2} + \phi_{1,2}y_{t-1} + \varepsilon_t & \text{om } y_{t-1} > c \end{cases} \quad (4.1)$$

vilket är en två regimers SETAR-modell av ordning ett. Notera att processen endast kan befinna i någon av dessa två regimer, och att regimskiften sker utan fördröjning. Dvs. den sk. regimskiftesfunktionen (eller övergångsfunktionen) är i detta fall stegfunktionen

Figur 4.1 Övergångsfunktion för TAR och SETAR-modeller



Notera också att t.ex. en $AR(p)$ -process endast består av en regim, och förblir därför linjär oavsett var i tidsserien vi befinner oss. Det är skiftena mellan dessa bitvis linjära modeller som gör SETAR sekvensen icke-linjär. I dessa modeller antas feltermerna ε_t vara o.i.f. $N(0, \sigma_\varepsilon^2)$, där storleken på σ_ε^2 förstås påverkar sannolikheten för regimskifte. Ett alternativt sätt att skriva SETAR-modellen (5.1) är

$$y_t = (\phi_{0,1} + \phi_{1,1}y_{t-1})(1 - I[y_{t-1} > c]) + (\phi_{0,2} + \phi_{1,2}y_{t-1})I[y_{t-1} > c] + \varepsilon_t \quad (4.2)$$

Där $I[A]$ är en indikatorfunktion med $I[A] = 1$ om händelse A inträffar, $I[A] = 0$ annars.

I en artikel från 1995 visar Simon Potter att icke-linjära modeller beskriver USA:s BNP kvartalsutveckling (1947:1-1990:4) betydligt bättre än linjära modeller. I artikeln jämförs en AR(5) -och en två regims TAR modells förmåga att beskriva serien $100[\log(GNP_t) - \log(GNP_{t-1})]$. Alla tänkbara mått (t , F , MSE, AIC, ...) gav i studien stöd för TAR-modellen.

I många makroekonomiska tidsserier sker dock regimskiftet gradvis och inte så plötsligt som TAR -och specialfallet SETAR-modeller föreslår. En logisk utveckling av dessa modeller är de sk. STAR (Smooth Transition Autoregressive)-modellerna där regimskiftet tillåts vara kontinuerligt. Ideén om mjuk övergång kom med Bacon och Watts (1971), men det var Chan och Tong (1986) som introducerade STAR-modellerna i tidsserielitteraturen. Indikatorfunktionen i (4.2) ersätts i STAR-modellerna med den kontinuerliga övergångsfunktionen $G(y_{t-1}; \gamma, c)$ vilket ger

$$y_t = (\phi_{0,1} + \phi_{1,1}y_{t-1})(1 - G(y_{t-1}; \gamma, c)) + (\phi_{0,2} + \phi_{1,2}y_{t-1})G(y_{t-1}; \gamma, c) + \varepsilon_t \quad (4.3)$$

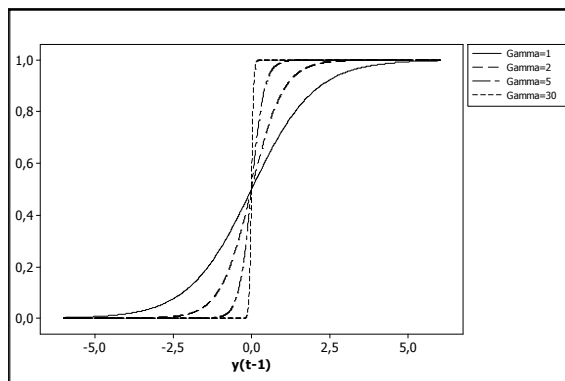
där $0 \leq G(y_{t-1}; \gamma, c) \leq 1$ och $G(y_{t-1}; \gamma, c) \rightarrow 1$ då $y_{t-1} \rightarrow \infty$. Utjämningsparametern γ indikerar hur snabbt övergången sker mellan regimerna i STAR-modellen.

En flitigt använd övergångsfunktion är den logistiska funktionen

$$G(y_{t-1}; \gamma, c) = \frac{1}{1 + e^{-\gamma(y_{t-1} - c)}} \quad (4.4)$$

Med övergångsfunktionen (4.4) insatt i (4.3) fås den sk. LSTAR (Logistic STAR)-modellen. I figur 4.2 visas den logistiska övergångsfunktionen för olika värden på γ , och där $c = 0$.

Figur 4.2 Övergångsfunktion för en LSTAR-modell för olika värden på γ



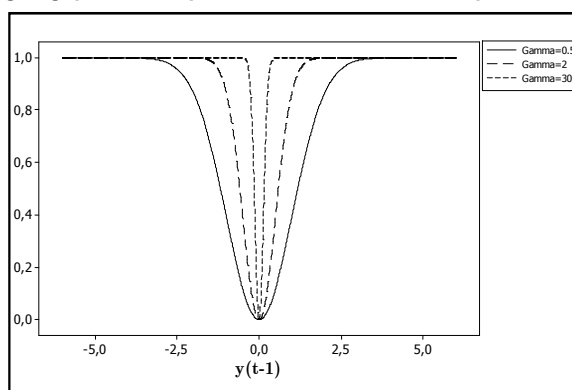
Vi ser tydligt att då γ är stort närmar sig den logistiska övergångsfunktionen $G(y_{t-1}; \gamma, c)$ indikatorfunktionen i figur 4.1. SETAR-modellen (4.2) kan alltså för stora värden på γ approximeras med LSTAR-modellen. I specialfallet då $\gamma = 0$ är LSTAR-modellen linjär ($G(y_{t-1}; \gamma, c)$ är konstant). Första ordningens autokorrelation, ρ_1 , i SETAR-modellen är lika med $\phi_{1,1}$ eller $\phi_{1,2}$ beroende på vilken regime som gäller. I STAR-modellen går ρ_1 gradvis från $\phi_{1,1}$ mot $\phi_{1,2}$ då y_{t-1} ökar.

En annan vanlig övergångsfunktion är den exponentiella funktionen

$$G(y_{t-1}; \gamma, c) = 1 - e^{-\gamma(y_{t-1}-c)^2} \quad (4.5)$$

Med övergångsfunktionen (4.5) insatt i (4.3) fås den sk. ESTAR (Exponential STAR)-modellen. I figur 4.3 visas den exponentiella övergångsfunktionen för olika värden på γ , med $c = 0$.

Figur 4.3 Övergångsfunktion för en ESTAR-modell för olika värden på γ



Då $\gamma \rightarrow 0$, eller om $\gamma \rightarrow \infty$, går ESTAR-modellen mot en linjär AR(p)-modell, eftersom $G(y_{t-1}; \gamma, c)$ då är konstant. ESTAR-modellen har i flera studier (t.ex. Michael, Nobay och Peel, (1997)) visat sig fungera speciellt bra under perioder nära vändpunkter, dvs. i perioder då y_{t-1}^2 är väldigt stor.

Icke-linjära modeller kan innehålla endogen dynamik, vilket innebär att y_t varierar även i avsaknandet av chocker (ε_t). Detta står i kontrast till de linjära modellerna där variationen endast orsakas av exogena chocker.

Även de enklaste SETAR, STAR och ESTAR-modellerna, där en AR(1)-modell använts i två regimer, är kapabla att generera en stor mängd dynamiska mönster. Givetvis ökar möjligheterna och flexibiliteten ytterligare om man tillåter för AR-modeller av högre ordning, eventuellt i fler än två regimer. T.ex kan en SETAR-modell i två regimer av ordning p_1 och p_2 skrivas

$$\begin{aligned} y_t = & (\phi_{0,1} + \phi_{1,1}y_{t-1} + \dots + \phi_{p_1,1}y_{t-p_1}) (1 - I[y_{t-1} > c]) \\ & + (\phi_{0,2} + \phi_{1,2}y_{t-1} + \dots + \phi_{p_2,2}y_{t-p_2}) I[y_{t-1} > c] + \varepsilon_t \end{aligned} \quad (4.6)$$

En STAR eller ESTAR-modell i två regimer av ordning p_1 och p_2 skrivs

$$\begin{aligned} y_t = & (\phi_{0,1} + \phi_{1,1}y_{t-1} + \dots + \phi_{p_1,1}y_{t-p_1}) (1 - G(y_{t-1}; \gamma, c)) \\ & + (\phi_{0,2} + \phi_{1,2}y_{t-1} + \dots + \phi_{p_2,2}y_{t-p_2}) G(y_{t-1}; \gamma, c) + \varepsilon_t \end{aligned} \quad (4.7)$$

En central och svår fråga är hur man bestämmer p_1 och p_2 i (4.6) och (4.7). En enkel metod som ofta används är att helt enkelt anpassa en linjär AR(p)-modell för y_t , och hoppas att ordningen, p (som förmodligen baseras på informationskriterier som AIC eller BIC), är lämplig att använda för båda regimerna i den icke-linjära modellen. Metoden ger dock osäkra resultat, inte sällan väljs p för lågt. Relativt enkla icke-linjära modeller ger upphov till komplicerade autokorrelationsstrukturer vilka bara kan beskrivas med AR modeller där p är väldigt stor.

Vid beräkandet av t.ex. AIC och BIC ges bestraffning över hela observationsmängden vid införandet av en ny parameter. Dessa mått är därför missvisande vid användandet på regimskiftesmodeller, eftersom regimerna kan innehålla olika många observationer. Tong (1990) föreslog att istället ge

bestraffning för nya parametrar bara för de observationer där de används. Dvs. AIC och BIC för en tvåregims TAR/SETAR-modell beräknas som

$$AIC(p_1, p_2) = n_1 \ln \hat{\sigma}_1^2 + n_2 \ln \hat{\sigma}_2^2 + 2(p_1 + 1) + 2(p_2 + 1)$$

och

$$BIC(p_1, p_2) = n_1 \ln \hat{\sigma}_1^2 + n_2 \ln \hat{\sigma}_2^2 + (p_1 + 1) \ln n_1 + (p_2 + 1) \ln n_2$$

där $\hat{\sigma}_i^2$ och n_i , $i = 1, 2$ är residualvariansen, respektive antalet observationer i regim i .

Stationaritetskraven för en tvåregims TAR/SETAR-modell är inte så restriktiva som man skulle kunna tro. T.ex. visade Chan et al. (1985) att en TAR och SETAR-modell med en AR(1)-modell i två regimer är stationära om en av följande villkor är uppfyllda

- (1) $\max(|\phi_{1,1}|, |\phi_{1,2}|) < 1$
- (2) $\phi_{1,1} = 1, \phi_{1,2} < 1, \phi_{0,1} > 0$
- (3) $\phi_{1,1} < 1, \phi_{1,2} = 1, \phi_{0,2} < 0$
- (4) $\phi_{1,1} = 1, \phi_{1,2} = 1, \phi_{0,2} < 0 < \phi_{0,1}$
- (5) $\phi_{1,1}\phi_{1,2} = 1, \phi_{1,1} < 0, \phi_{0,2} + \phi_{0,1}\phi_{1,2} > 0$

Där villkor (1) innebär att båda AR(1)-processerna är stationära, vilket alltså inte är ett nödvändigt villkor för stationaritet.

Maximal flexibilitet fås förstås om man förutom att tillåta olika ordning på AR modellerna, dessutom tillåter fler än två regimer. Vi kan t.ex. låta regimerna bestämmas av två variabler, vilket ger upphov till fler regimer. En fyrraregims SETAR-modell skulle kunna skrivas

$$y_t = \begin{cases} \phi_{0,1} + \phi_{1,1}y_{t-1} + \dots + \phi_{p_1,1}y_{t-p_1} + \varepsilon_t & \text{om } y_{t-1} \leq c_1 \text{ och } y_{t-2} \leq c_2 \\ \phi_{0,2} + \phi_{1,2}y_{t-1} + \dots + \phi_{p_2,2}y_{t-p_2} + \varepsilon_t & \text{om } y_{t-1} \leq c_1 \text{ och } y_{t-2} > c_2 \\ \phi_{0,3} + \phi_{1,3}y_{t-1} + \dots + \phi_{p_3,3}y_{t-p_3} + \varepsilon_t & \text{om } y_{t-1} > c_1 \text{ och } y_{t-2} \leq c_2 \\ \phi_{0,4} + \phi_{1,4}y_{t-1} + \dots + \phi_{p_4,4}y_{t-p_4} + \varepsilon_t & \text{om } y_{t-1} > c_1 \text{ och } y_{t-2} > c_2 \end{cases}$$

4.2 Modeller där regimerna bestäms av en icke-observerbar stokastisk process

I den andra klassen av regimskiftesmodeller antas att regimen vid tidpunkt t inte kan observeras, men att den kan bestämmas från den stokastiska processen s_t . Om vi har en process med två regimer kan s_t bara anta värdena 1 och 2. Med en AR(1)-modell i de två regimerna skrivs modellen

$$y_t = \begin{cases} \phi_{0,1} + \phi_{1,1}y_{t-1} + \varepsilon_t & \text{om } s_t = 1 \\ \phi_{0,2} + \phi_{1,2}y_{t-1} + \varepsilon_t & \text{om } s_t = 2 \end{cases} \quad (4.8)$$

Modellen (4.14) påminner mycket om en TAR-modell, den stora skillnaden är att vi här beräknar sannolikheter för ett regimskifte. Dessa modeller utvecklades av Hamilton (1989), och kallas för Markov switching-modeller (MSW). Skälet till namnet är att s_t antas följa en (1:a ordningens) Markovprocess. Övergångssannolikheterna definieras som

$$\begin{aligned} P(s_t = 1 | s_{t-1} = 1) &= p_{11} \\ P(s_t = 2 | s_{t-1} = 1) &= p_{12} \\ P(s_t = 1 | s_{t-1} = 2) &= p_{21} \\ P(s_t = 2 | s_{t-1} = 2) &= p_{22} \end{aligned}$$

där förstås $p_{12} = 1 - p_{11}$ och $p_{21} = 1 - p_{22}$. I MSW-modeller beror persistensen både på de autoregressiva parametrarna *och* på övergångssannolikheterna. T.ex. om $\phi_{1,1} > \phi_{1,2}$ och om p_{11} samtidigt är stor tenderar processen stanna kvar i regim 1 där persistensen är stor. Om dessutom p_{22} är liten blir effekten än större, dvs. processen kommer ha en tendens att fort lämna regim 2 till förmån för den resistent (och långlivade) regim 1. Den förväntade tiden som processen stannar i t.ex. regim 1 är $1/p_{11}$. Eftersom p_{11} och p_{22} är okända måste de skattas tillsammans med de autoregressiva parametrarna. Som med TAR-modeller gäller att om en regim inträffar sällan kommer dess parameterskattningar bli dåliga. Övergångssannolikheterna är alla betingade, t.ex. gäller att om processen är i regim 1 så är $1 - p_{11}$ den betingade sannolikheten att processen byter till regim 2.

Det är också av intresse att beräkna de obetingade sannolikheterna för att processen befinner sig i regim 1 (p_1) och regim 2 (p_2). Matrisen för övergångssannolikheterna kallas övergångsmatris ($\mathbf{P}$) och skrivs

$$\mathbf{P} = \begin{vmatrix} p_{11} & p_{21} \\ p_{12} & p_{22} \end{vmatrix} = \begin{vmatrix} p_{11} & 1 - p_{22} \\ 1 - p_{11} & p_{22} \end{vmatrix} \quad (4.9)$$

Egenvärdena för en N regims Markovkedja fås som lösningarna till $|\mathbf{P} - \mathbf{I}_N \lambda| = 0$. Här är $N = 2$ så vi får

$$\begin{vmatrix} p_{11} - \lambda & 1 - p_{22} \\ 1 - p_{11} & p_{22} - \lambda \end{vmatrix} = 0 \quad \Leftrightarrow \quad \dots (\lambda - 1) (\lambda + 1 - p_{11} - p_{22}) = 0$$

dvs. egenvärdena är $\lambda_1 = 1$ och $\lambda_2 = -1 + p_{11} + p_{22}$. Egenvektorn för λ_1 är

$$\begin{bmatrix} (1 - p_{22}) / 2 - p_{11} - p_{22} \\ (1 - p_{11}) / 2 - p_{11} - p_{22} \end{bmatrix}$$

Den obetingade sannolikheten att processen befinner sig i regim 1 en given tidpunkt är alltså

$$p_1 = \frac{1 - p_{22}}{2 - p_{11} - p_{22}}$$

För regim 2 är denna sannolikhet

$$p_2 = \frac{1 - p_{11}}{2 - p_{11} - p_{22}}$$

Egenvektorn för λ_2 är $\begin{bmatrix} -1 \\ 1 \end{bmatrix}$. Övergångsmatrisen kan dekomponeras som

$$\mathbf{P} = \mathbf{T} \mathbf{\Lambda} \mathbf{T}^{-1}$$

där $\mathbf{T}$ är den icke-singulära matrisen av övergångsmatrisens egenvektorer, och $\mathbf{\Lambda}$ är en diagonalmatris med övergångsmatrisens egenvärden på diagonalen. Det följer nu att t.ex.

$$\begin{aligned} \mathbf{P}^2 &= \mathbf{T} \mathbf{\Lambda} \mathbf{T}^{-1} \times \mathbf{T} \mathbf{\Lambda} \mathbf{T}^{-1} \\ &= \mathbf{T} \times \mathbf{\Lambda} \times (\mathbf{T}^{-1} \mathbf{T}) \times \mathbf{\Lambda} \times \mathbf{T}^{-1} \\ &= \mathbf{T} \times \mathbf{\Lambda} \times \mathbf{I} \times \mathbf{\Lambda} \times \mathbf{T}^{-1} = \\ &= \mathbf{T} \mathbf{\Lambda}^2 \mathbf{T}^{-1} \end{aligned}$$

och att

$$\begin{aligned} \mathbf{P}^m &= \mathbf{T} \times \mathbf{\Lambda} \times (\mathbf{T}^{-1} \mathbf{T}) \times \mathbf{\Lambda} \times (\mathbf{T}^{-1} \mathbf{T}) \times \dots (m \text{ gånger}) \dots \times \mathbf{\Lambda} \times \mathbf{T}^{-1} \\ &= \mathbf{T} \mathbf{\Lambda}^m \mathbf{T}^{-1} \end{aligned}$$

I vårt fall med en tvåregims markovkedja fås

$$\begin{aligned}\mathbf{P}^m &= \begin{bmatrix} \frac{1-p_{22}}{2-p_{11}-p_{22}} & -1 \\ \frac{1-p_{11}}{2-p_{11}-p_{22}} & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & \lambda^m \end{bmatrix} \begin{bmatrix} 1 & 1 \\ \frac{-(1-p_{11})}{2-p_{11}-p_{22}} & \frac{1-p_{22}}{2-p_{11}-p_{22}} \end{bmatrix} \\ &= \begin{bmatrix} \frac{(1-p_{22})-\lambda_2^m(1-p_{11})}{2-p_{11}-p_{22}} & \frac{(1-p_{22})-\lambda_2^m(1-p_{22})}{2-p_{11}-p_{22}} \\ \frac{(1-p_{11})-\lambda_2^m(1-p_{11})}{2-p_{11}-p_{22}} & \frac{(1-p_{11})-\lambda_2^m(1-p_{22})}{2-p_{11}-p_{22}} \end{bmatrix}\end{aligned}$$

Dvs. om processen just nu befinner sig i regim 2 så är sannolikheten att den m tidpunkter senare befinner sig i regim 1

$$\mathbf{P}(s_{t+m} = 1 | s_t = 2) = \frac{(1-p_{22}) - \lambda_2^m(1-p_{22})}{2-p_{11}-p_{22}}$$

där $\lambda_2 = -1 + p_{11} + p_{22}$.

4.3 Test för icke-linjäritet

En sökning efter en lämplig modell till en given tidsserie bör alltid inledas med ett test för icke-linjäritet. Man skiljer här på generella -och specifika test för icke-linjäritet. De generella testen används för att avgöra om en tidsserie är linjär eller icke-linjär, medan de specifika testen är utformade för att kunna avgöra formen av icke-linjäritet. Dessa två klasser av test behandlas separat i avsnitt 4.3.1 och 4.3.2.

4.3.1 Generella test för icke-linjäritet

Under nollhypotesen om linjäritet antas residualerna vara oberoende. Avvikelser från detta oberoende kan indikera på att den linjära modellen är olämplig. Det här är grandidén bakom många test för icke-linjäritet. De generella testen kan indelas i parametriska -och icke-parametriska test. Nedan presenteras några av de mest använda i dessa två klasser av test.

McLeod och Li (1983) använder Ljung-Box statistikan på de kvadrerade residualerna från en ARMA(p, q)-modell för att testa modellens lämplighet. Teststatistikan är

$$Q(m) = T(T+2) \sum_{i=1}^m \frac{\hat{\rho}_i^2(a_t^2)}{T-i} \quad (4.10)$$

där T är antalet observationer och m är ett lämpligt valt antal autokorrelationer som används i testet. a_t är residualserien och $\hat{\rho}_i^2(a_t^2)$ är autokorrelationsfunktionen av a_t^2 . Under nollhypotesen är $Q(m)$ asymptotiskt χ^2 -fördelad med $m - p - q$ frihetsgrader.

Ett annat icke-parametriskt test för icke-linjäritet är det s.k. BDS-testet, föreslaget av Brock *et al.* (1996). Testet undersöker avståndet mellan olika par av residualer. Låt w vara ett på förhand givet värde och låt ε_t och ε_{t-1} vara två realiseringar av sekvensen $\{\varepsilon_t\}$. Om $\{\varepsilon_t\}$ är oberoende kommer sannolikheten att avståndet mellan paret $(\varepsilon_i, \varepsilon_j)$ är mindre än w vara samma för alla i och j . Detta test fungerar dock dåligt på korta tidsserier, varför man ofta använder bootstraptekniker för att få fram kritiska värden.

Det mest kända parametriska testet för icke-linjäritet är RESET-testet (Regression Error Specification Test). Detta test som utformades av Ramsey (1969) utgår ifrån att om residualerna från den linjära modellen är oberoende ska de inte vara korrelerade med regressorerna i den skattade modellen eller med de skattade värdena. Dvs. en regression på residualerna ska inte vara signifikant förklarande. Testet utförs i två steg: först skattas den bästa tänkbara linjära modellen; låt $\{e_t\}$ vara den linjära modellens residualer och $\hat{y}_t$ dess skattade värden. Därefter estimeras för ett på förhand valt värde H regressionsekvationen

$$e_t = \delta \mathbf{z}_t + \sum_{h=2}^H \alpha_h \hat{y}_t^h \quad (4.11)$$

där $\mathbf{z}_t$ är vektorn av variabler i den skattade modellen. T.ex. om den linjära ARMA(2,1)-modell används för att beskriva y_t , så består $\mathbf{z}_t$ av en eventuell konstant, y_{t-1} , y_{t-2} och ε_{t-1} . Testet kan också tillämpas på regressionsmodeller eftersom $\mathbf{z}_t$ kan bestå av exogena förklarande variabler. Under nollhypotesen om linjäritet ($H_0 : \alpha_1 = \alpha_2 = \dots = \alpha_H$) är (4.11) asymptotiskt F -fördelad. Om tidsserien y_t är linjär kommer regressionen (4.11) ha liten förklaringsgrad, F -statistikan kommer vara liten och nollhypotesen accepteras. RESET-testet är enkelt att utföra och har relativt bra styrka att avslöja icke-linjäritet.

4.3.2 Specifika test för icke-linjäritet

Specifika test för icke-linjäritet har förmåga att avgöra den icke-linjära formen för en given tidsserie. Man låter den linjära modellen formulera nollhypotesen och olika former av icke-linjäritet (TAR, SETAR, STAR, MSW,

...) stå som mothypotes. Dessa olika test lider av samma problem, nämligen att de icke-linjära modellerna innehåller parametrar som inte existerar i de linjära modellerna under nollhypotesen. Test vid test av TAR-modell är inte tröskelvärde c definierad under nollhypotesen, i test av STAR-modeller är dessutom inte utjämningsparametern γ definierad. Vid test av MSW-modeller är inte övergångssannolikheterna p_{11} och p_{22} definierade under nollhypotesen. Konsekvensen av detta är att statistisk teori inte längre kan användas för att ta fram den asymptotiska fördelningen för dessa teststatistikor. Teststatistikorna har istället en icke-standard fördelning för vilka de kritiska värdena bestäms med simuleringsmetoder. Nedan behandlas de olika testen med början vid test av TAR/SETAR-modeller.

Den kanske viktigaste frågan är om de extra regimerna i regimskiftesmodeller signifikant ökar förklaringsgraden av serien y_t i förhållande till linjära (enregims) AR-modeller. Hypotesen av intresse är här $H_0 : \phi_{i,1} = \phi_{i,2}$ i modell (4.6). Problemet är att tröskelvärde c är oidentifierat under H_0 . En lösning är att använda skattningarna i TAR/SETAR-modeller och definiera F -statistikan

$$F_n(\hat{c}) = n \left(\frac{\tilde{\sigma}_n^2 - \hat{\sigma}_n^2}{\hat{\sigma}_n^2} \right) \quad (4.12)$$

där $\tilde{\sigma}_n^2$ är skattningen av residualvariansen under nollhypotesen om linjäritet, dvs.

$$\tilde{\sigma}_n^2 = \frac{1}{n} \sum_{t=1}^n \left(y_t - \mathbf{x}_t' \hat{\phi}_{i,1} \right)^2$$

och där

$$\hat{\phi}_{i,1} = \left(\sum_{t=1}^n \mathbf{x}_t \mathbf{x}_t' \right)^{-1} \left(\sum_{t=1}^n \mathbf{x}_t y_t \right)$$

$\hat{\phi}_{i,1}$ är OLS-skattningen av $\phi_{i,1}$ under antagandet om att $\phi_{i,1} = \phi_{i,2}$. $F_n(\hat{c})$ i ekvation (4.12) är en monoton funktion av $\hat{\sigma}_n^2$. Dvs. $F_n(\hat{c})$ ökar alltid då $\hat{\sigma}_n^2$ minskar och vice versa. Eftersom $\hat{c}$ minimerar residualvariansen är $F_n(\hat{c})$ ekvivalent med

$$F_n(\hat{c}) = \sup F_n(c)$$

där

$$F_n(c) = n \left(\frac{\tilde{\sigma}_n^2 - \hat{\sigma}_n^2(c)}{\hat{\sigma}_n^2(c)} \right)$$

$F_n(c)$ kan också beräknas som nR^2 , dvs. $F_n(c)$ följer asymptotiskt en χ^2 -fördelning med $p + 1$ frihetsgrader.

Även då vi testar linjäritet mot en STAR-modell används lämpligen en Lagrange-Multiplikator (LM) med asymptotisk χ^2 -fördelning under nollhypotesen. En fördel med LM test är att vi inte behöver skatta modellen under mothypotesen (STAR). Hypoteserna kan här formuleras på olika ekvivalenta sätt. Förutom samma AR-parametrar i de två regimerna ($H_0 : \phi_{i,1} = \phi_{i,2}$), kan nollhypotesen om linearitet uttryckas $H'_0 : \gamma = 0$. Om $\gamma = 0$ blir t.ex. den logistiska övergångsfunktionen (4.4) lika med 0.5 för alla värden på y_{t-1} , och STAR-modellen reduceras till en linjär AR modell med parametrarna $(\phi_{i,1} + \phi_{i,2})/2$. Luukkonen, Saikkonen och Teräsvirta (1988) visade att nollhypotesen $H'_0 : \gamma = 0$ i STAR-modellen (4.3) är ekvivalent med att testa $H''_0 : \beta_1 = 0$ i regressionsmodellen

$$y_t = \beta_{0,0} + \beta'_0 \tilde{\mathbf{x}}_t + \beta'_1 \tilde{\mathbf{x}}_t y_{t-1} + \eta_t$$

där $\tilde{\mathbf{x}}_t = (y_{t-1}, \dots, y_{t-p})'$ och $\beta_j = (\beta_{1,j}, \dots, \beta_{p,j})'$, $j = 0, 1$.

Teststatistikan för nollhypotesen $H''_0 : \beta_1 = 0$ är asymptotiskt χ^2 -fördelad med p frihetsgrader.

Ett sätt att testa relevansen av en MSW-modell är med Likelihood kvottest (LK) med vilken nollhypotesen om linjäritet ($H_0 : \phi_1 = \phi_2$) testas mot en MSW-modell med teststatistikan

$$LK_{MSW} = \mathcal{L}_{MSW} - \mathcal{L}_{AR} \quad (4.13)$$

där $\mathcal{L}_{MSW}$ och $\mathcal{L}_{AR}$ är värdena på MSW-modellens respektive AR-modellens log likelihoodfunktioner. Hansen (1992) visade att (4.13) har en icke-standard fördelning där de kritiska värdena bestäms med simulering. Denna dator-intensiva simuleringsprocedur påbörjas med att generera ett stort antal serier, y_t^* , från modellen som gäller under nollhypotesen. Därefter estimeras både AR-modeller och MSW-modeller för varje sådan genererad serie, och LK_{MSW} beräknas enligt (4.13). Dessa teststatistikor används därefter för att skatta fördelningen för teststatistikan under H_0 . Genom att beräkna andelen genererade serier för vilka LK_{MSW}^* överstiger LK_{MSW} för den observerade tidsserien erhålls p-värden för LK-statistikan.

4.4 Estimering

En tvåregims TAR-modell kan med indikatorfunktion skrivas

$$y_t = (\phi_{0,1} + \phi_{1,1}y_{t-1} + \dots + \phi_{p,1}y_{t-p}) I(q_{t-1} \leq c) + (\phi_{0,2} + \phi_{1,2}y_{t-1} + \dots + \phi_{p,2}y_{t-p}) I[q_{t-1} > c] + \varepsilon_t \quad (4.14)$$

där $q_{t-1} = q(y_{t-1}, \dots, y_{t-p})$ är en känd funktion. För att underlätta kommande uttryck låter vi

$$\mathbf{x}_t = (1 \quad y_{t-1} \quad \dots \quad y_{t-p})'$$

och

$$\mathbf{x}_t(c) = (\mathbf{x}_t' I(q_{t-1} \leq c) \quad \mathbf{x}_t' I(q_{t-1} > c))'$$

Ekvation (4.14) kan nu skrivas som

$$y_t = \mathbf{x}_t' \boldsymbol{\phi}_{p,1} I(q_{t-1} \leq c) + \mathbf{x}_t' \boldsymbol{\phi}_{p,2} I(q_{t-1} > c) + \varepsilon_t \quad (4.15)$$

där $\boldsymbol{\phi}_{p,1}' = (\phi_{0,1} \quad \phi_{1,1} \quad \dots \quad \phi_{p,1})$ och $\boldsymbol{\phi}_{p,2}' = (\phi_{0,2} \quad \phi_{1,2} \quad \dots \quad \phi_{p,2})$. Ekvation (5.9) kan också uttryckas som

$$y_t = \mathbf{x}_t(c)' \boldsymbol{\theta} + \varepsilon_t \quad (4.16)$$

där $\boldsymbol{\theta} = (\boldsymbol{\phi}_{p,1}' \quad \boldsymbol{\phi}_{p,2}')'$. Parametrarna av intresse är här $\boldsymbol{\theta}$ och c som (eftersom (4.16) är en regressionsekvation) lämpligen estimeras med minsta kvadratmetoden (MK). Under antagandet att ε_t är o.i.f. $N(0, \sigma^2)$ är MK-skattningar ekvivalenta med maximum likelihoodskattningar (ML). För ett givet värde på c fås skattningarna av $\boldsymbol{\theta}$ som

$$\hat{\boldsymbol{\theta}}(c) = \left(\sum_{t=1}^n \mathbf{x}_t(c) \mathbf{x}_t(c)' \right)^{-1} \left(\sum_{t=1}^n \mathbf{x}_t(c) y_t \right)$$

där $\hat{\mathbf{e}}_t(c) = y_t - \mathbf{x}_t(c)' \hat{\boldsymbol{\theta}}(c)$ med residualvariansen

$$\hat{\sigma}_n^2(c) = \frac{1}{n} \sum_{t=1}^n \hat{\mathbf{e}}_t(c)^2 \quad (4.17)$$

MK-skattningen av c minimerar (4.17):

$$\hat{c} = \arg \min \hat{\sigma}_n^2(c) \quad (4.18)$$

För att finna $\hat{c}$ i ekvation (4.18) utförs MK-regressioner av (4.16), där c sätts lika med q_{t-1} för alla q_{t-1} . För varje regression beräknas residualvariansen $\hat{\sigma}_n^2(c)$. Slutligen väljs det värde på $\hat{c}$ som minimerar denna varians. Det här kan också uttryckas som

$$\hat{c} = \arg \min \hat{\sigma}_n^2(q_{t-1}) \quad (4.19)$$

MK-skattningarna av θ ges av $\hat{\theta} = \hat{\theta}(\hat{c})$, med residualer $\hat{e}_t = y_t - \mathbf{x}_t(c)' \hat{\theta}$ och varians $\hat{\sigma}_n^2 = \hat{\sigma}_n^2(\hat{c})$.

I SETAR-modeller där $q_{t-1} = y_{t-d}$ måste i praktiken även parametern d skattas. Det går att med MK-metoden skatta d tillsammans med de andra parametrarna. Minimeringsproblemet i (4.19) måste nu modifieras för att inkludera sökningen för d . Så istället för att utföra n regressioner omfattar sökningen istället approximativt $\bar{d} \times n$ regressioner.

Chan (1993) visade att MK-skattningarna $\hat{c}$ är konsistenta och asymptotiskt oberoende av de andra skattningarna. Även skattningarna av de autoregressiva parametrarna är konsistenta och dessutom asymptotiskt normalfördelade.

Estimeringar av parametrarna i STAR-modellen (4.3) görs genom tillämpning av icke-linjära MK-skattningar. Dvs. $\theta = (\phi'_{p,1}, \phi'_{p,2}, \gamma, c)'$ skattas med

$$\hat{\theta} = \arg \min \sum_{t=1}^n [y_t - F(\mathbf{x}_t; \theta)]^2$$

där $F(\mathbf{x}_t; \theta) = \phi'_{p,1} \mathbf{x}_t (1 - G(y_{t-1}; \gamma, c)) + \phi'_{p,2} \mathbf{x}_t G(y_{t-1}; \gamma, c)$. Under antagandet att ε_t är normalfördelade är de icke-linjära MK-skattningarna ekvivalenta med ML-skattningar. Under vissa regularitetsvillkor är de icke-linjära MK-skattningarna konsistenta och asymptotiskt normalfördelade, dvs.

$$\sqrt{n} (\hat{\theta} - \theta_0) \rightarrow N(0, C)$$

där θ_0 är de sanna parametervärdena, och C är den asymptotiska kovariansmatrisen för $\hat{\theta}$. Estimeringen sker med någon icke-linjär optimeringsteknik (t.ex. Newton-Raphson).

En konsekvens av att regimerna inte är observerbara är att skattningar av MSW-modeller försvåras. Målet är förutom att skatta de autoregressiva parametrarna och övergångssannolikheterna också att skatta vilken regim som gäller vid varje tidpunkt i serien.

Givet MSW-modellen

$$y_t = \begin{cases} \phi_{0,1} + \phi_{1,1}y_{t-1} + \dots + \phi_{p,1}y_{t-p} + \varepsilon_t & \text{om } s_t = 1 \\ \phi_{0,2} + \phi_{1,2}y_{t-1} + \dots + \phi_{p,2}y_{t-p} + \varepsilon_t & \text{om } s_t = 2 \end{cases} \quad (4.20)$$

och att $\varepsilon_t \sim \text{o.i.f.} N$. Då gäller att tätheten för y_t betingat på s_t och på historiken Ω_{t-1} , är normalfördelad med väntevärde $\phi_{0,s_t} + \phi_{1,s_t}y_{t-1} + \dots + \phi_{p,s_t}y_{t-p}$ och varians σ^2 .

$$f(y_t | s_t = j, \Omega_{t-1}; \boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(y_t - \phi'_j \mathbf{x}_t)^2 / (2\sigma^2)} \quad (4.21)$$

där $\mathbf{x}_t = (1, y_{t-1}, \dots, y_{t-p})'$, $\phi_j = (\phi_{0,j}, \phi_{1,j}, \dots, \phi_{p,j})'$ för $j = 1, 2$. $\boldsymbol{\theta}$ är en vektor bestående av alla parametrar i modellen, $\boldsymbol{\theta} = (\phi'_1, \phi'_2, p_{11}, p_{22}, \sigma^2)$. Eftersom regimerna s_t inte går att observera fås den betingade log likelihoodfunktionen, $l_t(\boldsymbol{\theta})$, genom att betinga y_t endast på historiken Ω_{t-1} , dvs. $l_t(\boldsymbol{\theta}) = \ln f(y_t | \Omega_{t-1}; \boldsymbol{\theta})$. Täthetsfunktionen $f(y_t | \Omega_{t-1}; \boldsymbol{\theta})$ fås som

$$\begin{aligned} f(y_t | \Omega_{t-1}; \boldsymbol{\theta}) &= f(y_t, s_t = 1 | \Omega_{t-1}; \boldsymbol{\theta}) + f(y_t, s_t = 2 | \Omega_{t-1}; \boldsymbol{\theta}) \\ &= \sum_{j=1}^2 f(y_t | s_t = j, \Omega_{t-1}; \boldsymbol{\theta}) \times \mathbf{P}(s_t = j | \Omega_{t-1}; \boldsymbol{\theta}) \end{aligned} \quad (4.22)$$

Vi måste alltså för att kunna beräkna tätheten (4.22) skatta de betingade sannolikheterna för de olika regimerna givet processens historik, $\mathbf{P}(s_t = j | \Omega_{t-1}; \boldsymbol{\theta})$. Det visar sig att för att kunna beräkna ML-skattningar av modellens parametrar så måste vi för varje regim skatta tre olika sannolikheter. Dessa tre skattningar av regimsannolikheterna kallas för prognos, inferens och utjämnad inferens.

Om regimen vid tidpunkt $t-1$ är känd och inkluderad i informationsmängden Ω_{t-1} , skulle de s.k. optimala prognoserna av regimsannolikheterna vara lika med övergångssannolikheterna för s_t

$$\hat{\xi}_{t|t-1} = \mathbf{P}\xi_{t-1} \quad (4.23)$$

där

$$\widehat{\xi}_{t|t-1} = \begin{bmatrix} \mathbf{P}(s_t = 1 | \Omega_{t-1}; \boldsymbol{\theta}) \\ \mathbf{P}(s_t = 2 | \Omega_{t-1}; \boldsymbol{\theta}) \end{bmatrix}$$

och där $\xi_{t-1} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ om $s_t = 1$, och $\xi_{t-1} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ om $s_t = 2$. Matrisen $\mathbf{P}$ består av övergångssannolikheterna i (4.9).

Regimen vid tidpunkt $t - 1$ är dock i regel okänd. Det bästa man kan göra är att ersätta ξ_{t-1} i (4.23) med en skattning av sannolikheterna för varje regim vid $t - 1$ betingat på Ω_{t-1} . Den s.k. optimala inferensen av regimsannolikheterna är vektorn $\widehat{\xi}_{t-1|t-1}$. Givet ett startvärde $\widehat{\xi}_{1|0}$ och värden på parametrarna i $\boldsymbol{\theta}$ så går det att beräkna den optimala prognosen och inferensen för regimssannolikheterna genom att iterera på de två ekvationerna

$$\begin{aligned} \widehat{\xi}_{t|t} &= \frac{\widehat{\xi}_{t|t-1} \odot \mathbf{f}_t}{\mathbf{1}'(\widehat{\xi}_{t|t-1} \odot \mathbf{f}_t)} & (\text{inferens}) \\ \widehat{\xi}_{t+1|t} &= \mathbf{P}\widehat{\xi}_{t|t} & (\text{prognos}) \end{aligned}$$

där $\mathbf{f}_t$ är en vektor bestående av de betingade täthetsfunktionerna i (4.21)

för de två regimerna och $\mathbf{1} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$.

Startvärdena kan antingen skattas tillsammans med de andra parametrarna, eller också ses på som en vektor av konstanter som summerar till ett.

Om $\widehat{\xi}_{t|n}$ är beteckningen för den vektor som innehåller den s.k. utjämnade inferensen av regimsannolikheterna, dvs. skattningarna av sannolikheten att t.ex. regim j inträffar vid tidpunkt t , givet all tillgänglig information, $\mathbf{P}(s_t = j | \Omega_n; \boldsymbol{\theta})$. Dessa regimsannolikheter beräknas som

$$\widehat{\xi}_{t|n} = \widehat{\xi}_{t|t} \odot \left(\mathbf{P}' \left[\widehat{\xi}_{t+1|n} \div \widehat{\xi}_{t+1|t} \right] \right) \quad (4.24)$$

Algoritmen går baklänges genom serien, med start i $\widehat{\xi}_{n|n}$. Det visar sig att ML-skattningarna av övergångssannolikheterna är

$$\widehat{p}_{ij} = \frac{\sum_{t=2}^n \mathbf{P}(s_t = j, s_{t-1} = i | \Omega_n; \widehat{\boldsymbol{\theta}})}{\sum_{t=2}^n \mathbf{P}(s_{t-1} = i | \Omega_n; \widehat{\boldsymbol{\theta}})} \quad (4.25)$$

där $\widehat{\boldsymbol{\theta}}$ är ML-skattningen av $\boldsymbol{\theta}$. Det går också att visa att

$$\sum_{t=1}^n (y_t - \widehat{\phi}'_j \mathbf{x}_t) \mathbf{P}(s_t = j | \Omega_n; \widehat{\boldsymbol{\theta}}) = 0 \quad (4.26)$$

och

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{t=1}^n \sum_{j=1}^2 (y_t - \hat{\phi}'_j \mathbf{x}_t)^2 \mathbf{P}(s_t = j | \Omega_n; \hat{\boldsymbol{\theta}}) \quad (4.27)$$

I ekvation (4.26) är $\hat{\phi}_j$ en viktad minsta kvadrat skattning av y_t på x_t , med vikter givna av sannolikheten att regim j inträffar. Skattningarna av $\hat{\phi}_j$ kan alltså fås som

$$\hat{\phi}_j = \left(\sum_{t=1}^n \mathbf{x}_t(j) \mathbf{x}_t(j)' \right)^{-1} \left(\sum_{t=1}^n \mathbf{x}_t(j) y_t(j)' \right) \quad (4.28)$$

där

$$\begin{aligned} \mathbf{x}_t(j) &= \mathbf{x}_t \sqrt{\mathbf{P}(s_t = j | \Omega_n; \hat{\boldsymbol{\theta}})} \\ y_t(j) &= y_t \sqrt{\mathbf{P}(s_t = j | \Omega_n; \hat{\boldsymbol{\theta}})} \end{aligned}$$

Slutligen fås ML-skattningen av residualvariansen genom användande av (4.27) som medelvärdet av de kvadrerade residualerna från de två viktade minsta kvadrat regressionerna.

Denna snåriga skattningsprocedur är en tillämpning på den s.k. EM-algoritmen (Expectation maximization). I Hamilton (1990) ges en mer detaljerad framställning av denna algoritm och hur den fungerar.

Mc Culloch och Tsay (1993) använde istället MCMC metoder för att skatta MSW-modeller. De lät dessutom p_{11} och p_{22} vara logit eller probit funktioner av en förklarande variabel vid tidpunkt $t - 1$.

4.5 Diagnostik

Diagnosticeringen av icke-linjära modeller skiljer sig en del från diagnosticering av linjära modeller. T.ex. går det inte längre att tillämpa Ljung-Box för att testa residualernas oberoende. Däremot kan normalfördelningsantagandet testas som vanligt med t.ex. Anderson-Darling eller Jarque-Bera. Nedan redovisas hur de olika regimskiftesmodellerna diagnosticeras.

Vid diagnosticering av TAR/SETAR-modeller och STAR-modeller är tre test särskilt viktiga: test för residualernas oberoende, test för återstående icke-linjäritet och test för konstanta parametrar. Ett LM-test för ett p :te

ordningens seriellt beroende i ε_t kan fås som nR^2 , där R^2 är förklaringsgraden från regressionen av $\widehat{\varepsilon}_t$ på $\widehat{z}_t = \frac{\partial F(x_t; \theta)}{\partial \theta}$ och $\widehat{\varepsilon}_{t-1}, \widehat{\varepsilon}_{t-2}, \dots, \widehat{\varepsilon}_{t-p}$. Under nollhypotesen är teststatistikan asymptotiskt χ^2 -fördelad med p frihetsgrader.

Då vi har estimerat en icke-linjär modell så är en annan viktig fråga om denna modell har förmåga att förklara all icke-linjär dynamik i tidsserien. Ett sätt att undersöka detta är att test för återstående icke-linjäritet. För SETAR- och STAR-modeller är det naturligt att testa nollhypotesen om linjäritet mot alternativet att ytterligare regimer behövs för att fånga upp hela tidsseriens dynamik. T.ex. testas nollhypotesen att en tvåregims STAR-modell är lämplig för en given tidsserie mot mothypotesen att en tredje regim är nödvändig. Detta test är enkelt att utföra eftersom vi bara behöver skatta modellen under nollhypotesen. Däremot måste vi för att testa motsvarande i SETAR-modeller även skatta modellen i mothypotesen. Även här används LM-test med asymptotiskt χ^2 -fördelade testvariabler, se t.ex. Eitrheim och Teräsvirta (1996) för mer detaljer.

Ett specialfall av multipel-regimsmodellen

$$y_t = [\phi'_1 \mathbf{x}_t (1 - G_1(y_{t-1})) + \phi'_2 \mathbf{x}_t G_1(y_{t-1})] [1 - G_2(y_{t-2})] + [\phi'_3 \mathbf{x}_t (1 - G_1(y_{t-1})) + \phi'_4 \mathbf{x}_t G_1(y_{t-1})] G_2(y_{t-2}) + \varepsilon_t$$

uppstår om den andra övergångsfunktionen G_2 definieras för y_t istället för y_{t-2} . Vi får då en sk. tidsvarierande STAR-modell, vilken tillåter både icke-linjär dynamik och tidsvarierande parametrar. Genom att testa om $\gamma_2 = 0$ testas samtidigt för konstanta parametrar i tvåregims STAR-modellen (4.7) mot hypotesen om mjuk förändring i parametrarna. Detta test kan även användas för att testa för konstanta parametrar i SETAR-modeller. Detta görs genom att approximera SETAR-modellen med en STAR-modell där γ_1 sätts lika med ett stort värde.

I Hamilton (1996) behandlas hur MSW-modeller kan verifieras. Hamilton utvecklar test för autokorrelation i residualerna, heteroskedasticitet, felspecificering av Markovprocessen s_t samt för missade förklarade variabler. Dessa test är alla av LM-typ och har fördelen att vi bara behöver skatta modellen som gäller under nollhypotesen. I testen används scorestatistikan

$$h_t(\theta) = \frac{\partial \ln f(y_t | \Omega_{t-1}; \theta)}{\partial \theta}$$

Ett användningsområde för denna scorestatistika är att beräkna standardfel

för ML-skattningarna av θ . Dessa fås nämligen som

$$\sum_{t=1}^n h_t(\hat{\theta})h_t(\hat{\theta})'$$

Vi kan också använda scorestatistikan för att få fram LM-statistikor. Låt säga att vi vill testa om några variabler z_t saknas i en tvåregims MSW-modell. Vi testar då

$$y_t = \begin{cases} \phi_{0,1} + \phi_{1,1}y_{t-1} + \dots + \phi_{p,1}y_{t-p} + \varepsilon_t & \text{om } s_t = 1 \\ \phi_{0,2} + \phi_{1,2}y_{t-1} + \dots + \phi_{p,2}y_{t-p} + \varepsilon_t & \text{om } s_t = 2 \end{cases}$$

som ekvivalent kan skrivas

$$y_t = \phi_{0,s_t} + \phi_{1,s_t}y_{t-1} + \dots + \phi_{p,s_t}y_{t-p} + \varepsilon_t$$

mot

$$y_t = \phi_{0,s_t} + \phi_{1,s_t}y_{t-1} + \dots + \phi_{p,s_t}y_{t-p} + \delta'z_t + \varepsilon_t \quad (4.29)$$

dvs, vi testar i ekvation (4.29) hypotesen $H_0 : \delta' = 0$. Under denna nollhypotes blir scorestatistikan med avseende på δ

$$\left. \frac{\partial \ln f(y_t | \Omega_{t-1}; \theta)}{\partial \delta} \right|_{\delta=0} = \sum_{j=1}^2 (y_t - \hat{\phi}'_j \mathbf{x}_t) z_t \mathbf{P}(s_t = j | \Omega_n; \hat{\theta})$$

där $\hat{\theta}$ är ML-skattningar av parametervektorn $\theta' = (\phi'_1, \phi_2, p_{11}, p_{22}, \delta)$ under nollhypotesen. LM-testet för att testa H_0 ges av

$$n \left(\frac{1}{n} \sum_{t=1}^n h_t(\hat{\theta}) \right) \left(\frac{1}{n} \sum_{t=1}^n h_t(\hat{\theta})h_t(\hat{\theta})' \right)^{-1} \left(\frac{1}{n} \sum_{t=1}^n h_t(\hat{\theta}) \right)$$

vilken har en asymptotisk χ^2 -fördelning under nollhypotesen med antal frihetsgrader lika med antalet variabler i z_t .

4.6 Prognoser

Förutom att bättre förklara dynamiken i en viss tidsserie används förstås också icke-linjära modeller till prognoser. Att beräkna punktprognoser med icke-linjära modeller är mer komplicerat än med linjära modeller. För att se om det är mödan värt används ofta prognosfelen från en linjär modell som benchmark.

Ett sätt att beräkna prognoser av icke-linjära modeller är med parametrisk bootstrap. Låt T vara antalet observationer i tidsserien, och $\ell > 0$ vara prognoshorisonten. Dvs. det senast kända värdet är y_T och vi vill nu skapa prognoser för $y_{T+\ell}$. Med parametrisk bootstrap beräknas realiseringar av $y_{T+1}, \dots, y_{T+\ell}$ sekventiellt genom att först dra nya innovationer från modellens innovationsfördelning. Därefter beräknas $y_{T+\ell}$ med användande av modellen, datamaterialet samt prognoserna för $y_{T+1}, \dots, y_{T+\ell-1}$. Det här ger en realisering för $y_{T+\ell}$. Denna procedur upprepas M gånger så att vi får M realiseringar av $y_{T+\ell}$, $\left\{y_{T+\ell}^{(j)}\right\}_{j=1}^M$. Punktprognosen av $y_{T+\ell}$ fås sedan som medelvärdet av $y_{T+\ell}^{(j)}$. Dessa M realiseringar kan också användas för att få en empirisk fördelning för $y_{T+\ell}$.

Andra metoder för att beräkna punkt -och intervallprognoser med icke-linjära modeller behandlas i t.ex. Franses och van Dijk (2000), sid 118-125.

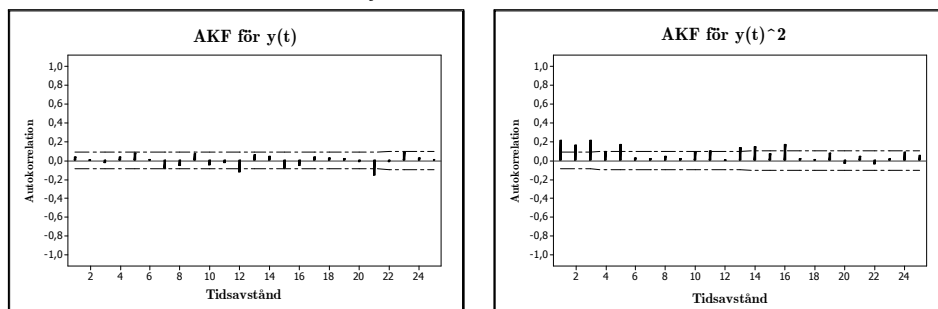
5 Icke-linjäritet i varians

I avsnitt 1.2 diskuterades det faktum att stora förändringar i många finansiella tidsserier kommer i kluster. Detta faktum observerades redan av Mandelbrot (1963) som trots det modellerade avkastningen som oberoende och identiskt fördelad över tiden. En annan ansats hade Engle (1982) som istället föreslog en modell som bygger på att slumptermerna är *beroende* av varandra. Engles nobelprisbelönta modell kallas ARCH (Autoregressive Conditional Heteroscedasticity) och i denna antas variansen i slumptermerna systematiskt bero på tidigare slumptermer. Om dessa ARCH-modeller och dess släktingar handlar detta kapitel. I delkapitel 5.1 beskrivs ARCH-modellen och några av dess närmaste linjära och icke-linjära släktingar (GARCH, IGARCH, FIGARCH, ...). I de följande avsnitten 5.2-5.5 avhandlas i tur och ordning test för heteroskedasticitet, estimering, diagnostik och prognoser. Kapitlet avslutas med en beskrivning av några alternativa metoder där man på olika sätt försöker filtrera bort heteroskedasticiteten, för att istället analysera den transformerade homoskedastiska serien.

5.1 Modeller för heteroskedastiska tidsserier

Innan vi ser på några olika modeller för heteroskedastiska tidsserier tittar vi återigen på den heteroskedastiska (diff log) OMX-S serien i figur 1.1. Om igen $y_t = 100(\ln P_t - \ln P_{t-1})$, där P_t är OMX-S slutkurs dag t under perioden 2005-07-01 till 2007-07-25 uttryckt i US dollar. Autokorrelationsfunktionen (AKF) för y_t och för y_t^2 redovisas i figur 5.1

Figur 5.1: AKF för y_t och y_t^2



Typiskt är att serien $\{y_t\}$ antingen är seriellt okorrelerad eller möjligen har svag autokorrelation, men att $\{y_t^2\}$ är beroende. Detta illustreras i figur 5.1 ovan där det verkar som om avkastningen är okorrelerad, men att det finns ett beroende i den kvadrerade avkastningen. Modeller för heteroskedastiska serier försöker på olika sätt att fånga detta beroende.

Det betingade medelvärdet och den betingade variansen av y_t givet Ω_{t-1} skrivs

$$\mu_t = E(y_t | \Omega_{t-1}), \quad \sigma_t^2 = Var(y_t | \Omega_{t-1}) = E[(y_t - \mu_t)^2 | \Omega_{t-1}] \quad (5.1)$$

Som vi såg i figur 5.1 är autokorrelationen liten i y_t (om den existerar överhuvudtaget). Anta att y_t följer en stationär ARMA(p, q)-modell, dvs.

$$y_t = \mu_t + a_t, \quad \text{där} \quad \mu_t = \phi_0 + \sum_{i=1}^p \phi_i y_{t-i} - \sum_{i=1}^q \theta_i a_{t-i} \quad (5.2)$$

Om (5.1) och (5.2) kombineras fås

$$\sigma_t^2 = Var(y_t | \Omega_{t-1}) = Var(a_t | \Omega_{t-1}) \quad (5.3)$$

Alla modeller i det här kapitlet försöker beskriva beteendet av σ_t^2 . Beroende

på hur σ_t^2 varierar används olika volatilitetsmodeller. Man kan här skilja på linjära -och icke-linjära volatilitetsmodeller. Den stora skillnaden är att linjära volatilitetsmodeller hanterar positiva och negativa chocker på samma sätt. Icke-linjära volatilitetsmodeller däremot skiljer på positiva och negativa chocker, vilket gör att de är bättre på att beskriva eventuella asymmetrier i den heteroskedastiska tidsserien. Dessa två klasser av modeller behandlas separat i avsnitt 5.1.1 och 5.1.2.

5.1.1 Linjära volatilitetsmodeller

Dessa modeller för betingad heteroskedasticitet kan indelas i två klasser där man i den första klassen använder en exakt funktion för att beskriva σ_t^2 . I den andra klassen används istället en stokastisk ekvation för att beskriva σ_t^2 . T.ex. tillhör ARCH -och GARCH-modeller den första klassen, medan de sk. stokastiska volatilitetsmodellerna tillhör den andra klassen.

ARCH-modellen

Engle (1982) var först med att försöka modellera heteroskedastiska tidsserier med sina ARCH-modeller. Idén med ARCH-modeller är dels att den medelvärdes-korrigerade avkastningen a_t är seriellt (linjärt) okorrelerad men beroende. Dessutom kan detta beroende förklaras med en enkel kvadratisk funktion av dess tidsförskjutna variabler. T.ex. antas i en ARCH(m)-modell att

$$a_t = \sigma_t \varepsilon_t, \quad \sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \dots + \alpha_m a_{t-m}^2 \quad (5.4)$$

där $\{\varepsilon_t\}$ är en sekvens av o.i.f stokastiska variabler med väntevärde 0 och varians 1, och där $\alpha_0 > 0, \alpha_i \geq 0$ för $i = 1, 2, \dots, m$. Koefficienterna α_i måste dessutom uppfylla vissa regularitetsvillkor för att variansen för a_t ska vara ändlig. I praktiken antas ofta att ε_t är standardiserat normalfördelad eller standardiserat t -fördelad. Vissa (inkl. Engle själv) föredrar att beteckna den betingade variansen i (5.4) h_t , och skriver alltså $a_t = \sqrt{h_t} \varepsilon_t$. Eftersom a_t till stor del förklaras av tidigare värden av serien a_t tenderar stora chocker att efterföljas av stora chocker och vice versa. ARCH-modellen har därför en tydlig intuitiv förmåga att beskriva heteroskedastiska serier. För att bättre förstå dessa ARCH-modeller studerar vi några egenskaperna hos ARCH(1)-modellen

$$a_t = \sigma_t \varepsilon_t, \quad \sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 \quad (5.5)$$

Det obetingade medelvärdet är

$$E(a_t) = E[E(a_t | \Omega_{t-1})] = E[\sigma_t E(\varepsilon_t)] = 0$$

Den obetingade variansen blir

$$Var(a_t) = E(a_t^2) = E[E(a_t^2 | \Omega_{t-1})] = E(\alpha_0 + \alpha_1 a_{t-1}^2) = \frac{\alpha_0}{1 - \alpha_1}$$

där $0 \leq \alpha_1 < 1$ för att variansen ska vara positiv. För att högre moment av fördelningen för a_t ska existera krävs ytterligare villkor på α_1 . För att t.ex. studera svansarna av a_t krävs att fjärde momentet (toppigheten) är ändligt. Under antagandet att ε_t i (5.4) är standardiserat normalfördelad (där toppigheten = 3) blir

$$E(a_t^4 | \Omega_{t-1}) = 3[E(a_t^2 | \Omega_{t-1})]^2 = 3(\alpha_0 + \alpha_1 a_{t-1}^2)^2$$

Därför blir

$$E(a_t^4) = E[E(a_t^4 | \Omega_{t-1})] = 3E(\alpha_0 + \alpha_1 a_{t-1}^2)^2 = 3E[\alpha_0^2 + 2\alpha_0\alpha_1 a_{t-1}^2 + \alpha_1^2 a_{t-1}^4]$$

Om a_t är stationär med toppighet $m_4 = E(a_t^4)$ fås

$$\begin{aligned} m_4 &= 3[\alpha_0^2 + 2\alpha_0\alpha_1 Var(a_t) + \alpha_1^2 m_4] \\ &= 3\alpha_0^2 \left(1 + 2\frac{\alpha_1}{1 - \alpha_1}\right) + 3\alpha_1^2 m_4 \end{aligned}$$

dvs

$$m_4 = \frac{3\alpha_0^2(1 + \alpha_1)}{(1 - \alpha_1)(1 - 3\alpha_1^2)}$$

Eftersom fjärde momentet av a_t är positivt måste $1 - 3\alpha_1^2 > 0$, dvs. $0 \leq \alpha_1 < 1/\sqrt{3}$. Villkoren blir mer komplicerade för ARCH-modeller av högre ordning. Den obetingade toppigheten för a_t är

$$\frac{E(a_t^4)}{[Var(a_t)]^2} = \kappa_a = \frac{3\alpha_0^2(1 + \alpha_1)}{(1 - \alpha_1)(1 - 3\alpha_1^2)} \times \frac{(1 - \alpha_1)^2}{\alpha_0^2} = 3\frac{1 - \alpha_1^2}{1 - 3\alpha_1^2} > 3$$

Toppigheten för a_t är alltså större än tre vilket innebär att fördelningen är mer toppig och har tjockare svansar än en standardiserad normalfördelning. Detta är en viktig egenskap eftersom vi från avsnitt 1.1 vet att stora förändringar i finansiella tidsserier är vanligare än förväntat under antagandet om normalfördelning.

Dynamiken i ARCH-modellen fångas upp i AKF. Låt

$$a_t^2 = \sigma_t^2 + (a_t^2 - \sigma_t^2)$$

Använd sedan (5.5) för att få

$$a_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + v_t \quad (5.6)$$

där $v_t = \sigma_t^2(\varepsilon_t^2 - 1)$ kan visas vara vitt brus. a_t^2 följer alltså en AR(1)-process med AKF

$$\rho_k = \alpha_1^k, \quad k = 0, 1, 2, \dots$$

De många fördelarna med ARCH-modeller är nämnda och kända, men det finns även några brister. Dels antas det i modellen att positiva och negativa chocker har samma effekt på volatiliteten - något som motsäger empirin (se figur 1.1 och 1.2). ARCH-modeller är dessutom väldigt restriktiva, t.ex. måste $0 \leq \alpha_1 < 1/\sqrt{3}$ för att fjärde momentet ska vara ändligt. Som vi såg ovan tenderar ARCH modeller att överskatta volatiliteten eftersom modellen reagerar långsamt på stora chocker (KOLLA UPP). Detta är en del av förklaringen till varför de generella ARCH-modellerna, även kallade GARCH (Generalized ARCH), är mer användbara i praktiken. Men den kanske viktigaste anledningen är att GARCH-modeller vanligen beskriver volatiliteten med färre parametrar.

GARCH-modellen

GARCH-modellerna som utvecklades av Bollerslev (1986) inkluderar även tidigare värden σ_{t-j}^2 i variansekvationen. Om $a_t = y_t - \mu_t$ är den medelvärdes-korrigerade logaritmerade avkastningen då följer a_t en GARCH(m, s)-modell om

$$a_t = \sigma_t \varepsilon_t, \quad \sigma_t^2 = \alpha_0 + \sum_{i=1}^m \alpha_i a_{t-i}^2 + \sum_{j=1}^s \beta_j \sigma_{t-j}^2 \quad (5.7)$$

där igen $\{\varepsilon_t\}$ är en sekvens av o.i.f stokastiska variabler med väntevärde 0 och varians 1. Vidare krävs att $\alpha_0 > 0, \alpha_i \geq 0, \beta_j \geq 0$ och $\sum_{i=1}^{\max(m,s)} (\alpha_i + \beta_i) < 1$, där förstås $\alpha_i = 0$ om $i > m$ och $\beta_i = 0$ om $j > s$. Det sista villkoret ($\sum_{i=1}^{\max(m,s)} (\alpha_i + \beta_i) < 1$) försäkrar oss om att den obetingade variansen är ändlig. Även i GARCH-modeller är det vanligt att anta att ε_t är standardiserat normal- eller t -fördelad. Givetvis reduceras modell (5.7) till en ARCH(m)-modell om $s = 0$.

Utvecklingen av ARCH-modeller till GARCH-modeller är den icke-linjära motsvarigheten till utvecklingen från AR-modeller till ARMA-modeller. För att se detta låter vi $\eta_t = a_t^2 + \sigma_t^2$ så att $\sigma_t^2 = a_t^2 - \eta_t$. Genom att sätta in $\sigma_{t-i}^2 = a_{t-i}^2 - \eta_{t-i}$, i (5.7) där $i = 0, 1, \dots, s$ ser vi att GARCH-modellen kan skrivas om som

$$a_t^2 = \alpha_0 + \sum_{i=1}^{\max(m,s)} (\alpha_i + \beta_i) a_{t-i}^2 + \eta_t - \sum_{j=1}^s \beta_j \eta_{t-j} \quad (5.8)$$

Ekvation (5.8) är en ARMA-form av a_t^2 . Den obetingade variansen fås på samma sätt som i ARMA-modeller som

$$E(a_t^2) = \frac{\alpha_0}{1 - \sum_{i=1}^{\max(m,s)} (\alpha_i + \beta_i)}$$

GARCH-modellens för- och nackdelar framkommer då man studerar GARCH(1, 1)-modellen

$$\sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \quad (5.9)$$

där $\alpha_0 > 0, \alpha_1 \geq 0, \beta_1 \geq 0$ och $(\alpha_1 + \beta_1) < 1$. Från (5.9) kan vi se att stora värden på a_{t-1}^2 och σ_{t-1}^2 ger stora värden på σ_t^2 . Det här innebär att ett stort värde på a_{t-1}^2 tenderar att efterföljas av ett stort a_t^2 . GARCH-modellen har alltså liksom ARCH-modellen förmåga att generera en heteroskedastisk serie. Det går att visa att om $1 - 2\alpha_1^2 - (\alpha_1 + \beta_1)^2 > 0$, då är

$$\frac{E(a_t^4)}{[E(a_t^2)]^2} = \kappa_a = \frac{3[1 - (\alpha_1 + \beta_1)^2]}{1 - 2\alpha_1^2 - (\alpha_1 + \beta_1)^2} > 3$$

I likhet med ARCH-modeller har alltså även GARCH-modeller tjockare

svansar än en standardiserad normalfördelning. AKF av a_t^2 i GARCH(1,1)-modellen kan visas vara

$$\rho_k = \begin{cases} \alpha_1 + \frac{\alpha_1^2 \beta_1}{1 - 2\alpha_1 \beta_1 - \beta_1^2} & \text{om } k = 1 \\ (\alpha_1 + \beta_1)^{k-1} \rho_1 & \text{om } k = 2, 3, \dots \end{cases}$$

Autokorrelationerna avtar alltså exponentiellt med faktorn $\alpha_1 + \beta_1$.

IGARCH-modellen

I många tillämpningar av GARCH(1,1)-modellen är summan av α_1 och β_1 väldigt nära ett. Då $\alpha_1 + \beta_1 = 1$ fås ett specialfall av GARCH-modellen som kallas IGARCH (Integrated GARCH). Anledningen till namnet är att om $\alpha_1 + \beta_1 = 1$ existerar en enhetsrot i ARMA(1,1)-modellen för a_t^2 i (5.8), som nu kan skrivas om som

$$(1 - B)a_t^2 = \alpha_0 + \eta_t - \beta_1 \eta_{t-1} \quad (5.10)$$

där B är bakåtoperatoren sådan att t.ex. $Ba_t^2 = a_{t-1}^2$. Den obetingade variansen, $E(a_t^2) = \alpha_0 / (1 - \alpha_1 - \beta_1)$, är uppenbarligen inte definierad för IGARCH(1,1)-modellen, modellen är alltså inte svagt stationär. IGARCH(1,1)-modellen kan dock vara strikt stationär, se Nelson (1990). Eftersom inte den obetingade variansen är definierad kan vi inte heller definiera AKF för a_t^2 . Ding och Granger (1996) visade emellertid hur approximativa autokorrelationer ges av

$$\rho_k \approx \frac{1}{3}(1 + 2\alpha_1)(1 + 2\alpha_1^2)^{-k/2} \quad (5.11)$$

Notera att autokorrelationerna i (5.11) avtar exponentiellt.

FIGARCH-modellen

I GARCH(1,1) avtar ibland AKF för a_t^2 för snabbt, detta oavsett hur låga värden uttrycket $1 - \alpha_1 - \beta_1$ antar. I dessa fall föreslog Baillie et al. (1996) en GARCH-modell med långt minne, FIGARCH (Fractionally Integrated GARCH). T.ex. fås FIGARCH(1, d , 0)-modellen enklast genom att i (5.10) introducera potensen d till differensoperatoren

$$(1 - B)^d a_t^2 = \alpha_0 + \eta_t - \beta_1 \eta_{t-1}$$

där $0 < d < 1$. Man kan visa att

$$\rho_k \approx [(1 - \beta_1)\Gamma(d)^{-1}] k^{d-1} \quad (5.12)$$

där $\Gamma(\cdot)$ är gammafunktionen. AKF (5.12) avtar hyperboliskt då k ökar. Denna modell har tydliga kopplingar till ARIMA-modeller med långt minne, de sk. ARFIMA(AutoRegressive Fractionally Integrated Moving Average)-modellerna. Gemensamt för både FIGARCH -och ARFIMA-modellerna är att $0 < d < 1$ (förutsatt att processen är stationär). En sådan fraktionellt integrerad process sägs ha långt minne vilket innebär att chocker i serien är väldigt persistenta (AKF avtar mycket långsamt). Ding och Granger (1996) visade att stickprovs AKF inledningsvis avtar snabbare än exponentiellt, men att den avtar (mycket) långsammare på längre tidsavstånd. Det här mönstret indikerar att volatiliteten kan bestå av olika komponenter där vissa har stor effekt på kort sikt för att sedan snabbt avta, medan andra komponenter är mer persistenta. Av denna anledning föreslår ibid. den sk. komponent GARCH-modellen som består av en viktad summa av två komponenter; en IGARCH -och en GARCH-modell.

GARCH-M modellen

Många finansiella teorier behandlar förhållandet mellan avkastningen och risken för en viss tillgång. Av denna anledning skapade Engle et al. (1987) den sk. GARCH-M (GARCH in Mean)-modellen som är designad för att kunna beskriva detta förhållande. Detta åstadkoms genom att inkludera en funktion av den betingade variansen σ_t^2 i modellen för det betingade medelvärde, t.ex.

$$y_t = \phi_0 + \phi_1 y_{t-1} + \delta g(\sigma_t^2) + a_t \quad (5.13)$$

där y_t är variabeln av intresse och där $g(\sigma_t^2)$ är en funktion av den betingade variansen av a_t . Vidare antas σ_t^2 följa en ARCH -eller en GARCH-process. I praktiken gör man det ofta enkelt för sig genom att låta $g(\sigma_t^2)$ vara en enhetsfunktion eller en kvadratrotsfunktion, dvs. $g(\sigma_t^2) = \sigma_t^2$ eller $g(\sigma_t^2) = \sigma_t$. δ kallas för riskparametern. Om $\delta > 0$ är avkastningen positivt korrelerad med tidigare volatilitet och vice versa. För att få en känsla för GARCH-M modellens egenskaper ser vi på modellen

$$y_t = \delta \sigma_t^2 + a_t \quad (5.14)$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 \quad (5.15)$$

vilket är (5.13) med $\phi_1 = 0$ och $g(\sigma_t^2) = \sigma_t^2$ där σ_t^2 följer en ARCH(1)-process. Om vi nu sätter in (5.15) i (5.14) fås

$$y_t = \delta\alpha_0 + \delta\alpha_1 a_{t-1}^2 + a_t$$

med det betingade väntevärdet

$$E(y_t) = \delta\alpha_0 \left(1 + \frac{\alpha_1}{1 - \alpha_1} \right)$$

där vi använt att $E(a_{t-1}^2) = \alpha_0/(1 - \alpha_1)$. På samma sätt går det att visa att den obetingade variansen av y_t är

$$\sigma_y^2 = \frac{\alpha_0}{1 - \alpha_1} + \frac{(\delta\alpha_1)^2 2\alpha_0^2}{(1 - \alpha_1)^2 (1 - 3\alpha_1^2)} \quad (5.16)$$

Eftersom den andra termen i högerledet i (5.16) är större än 0 är σ_y^2 alltid större än den obetingade variansen för y_t utan GARCH-M effekten. AKF för GARCH-M modellen i (5.14) och (5.15) är

$$\rho_k = \begin{cases} \frac{2\alpha_1^3 \delta^2 \alpha_0}{2\alpha_1^2 \delta^2 \alpha_0 + (1 - \alpha_1)(1 - 3\alpha_1^2)} & \text{om } k = 1 \\ \alpha_1^{k-1} \rho_1 & \text{om } k = 2, 3, \dots \end{cases}$$

Den stokastiska volatilitetsmodellen

En annan typ av modeller fås om man inkluderar chocker i den betingade variansen för a_t . Vi får då en sk. stokastisk volatilitetsmodell (SV). Denna modell som formulerades av Taylor (1986) skrivs

$$\begin{aligned} a_t &= \sigma_t \varepsilon_t \\ (1 - \alpha_1 B - \dots - \alpha_m B^m) \ln(\sigma_t^2) &= \alpha_0 + v_t \end{aligned} \quad (5.17)$$

där ε_t är o.i.f $N(0, 1)$, $v_t \sim$ o.i.f. $N(0, \sigma_v^2)$ och där $\{\varepsilon_t\}$ och $\{v_t\}$ är oberoende. Introducerandet av chocken v_t i (5.17) ökar modellens flexibilitet att förklara σ_t^2 , men också svårigheten att skatta modellens parametrar. Parametrarna i GARCH-modellen skattas genom standardtillämpning av ML-tekniker, se

avsnitt 5.3. Det som försvårar skattningen av SV-modeller är att σ_t^2 inte är observerbar, och kan heller inte rekonstrueras av historiken Ω_{t-1} . Det här innebär att ML-tekniker nu inte är tillämpbara på SV-modeller. Olika (simuleringsbaserade) skattningstekniker har föreslagits, se t.ex. Melino och Turnbull (1990).

Det går att visa att SV-modellen har en toppighet större än tre samt att AKF är

$$\rho_k = \frac{e^{\sigma_t^2 \gamma_1^k} - 1}{3e^{\sigma_t^2} - 1} \approx \frac{e^{\sigma_t^2} - 1}{3e^{\sigma_t^2} - 1} \gamma_1^k \quad (5.18)$$

där approximationen i (5.18) är god för små värden på σ_t^2 och/eller för stora värden på kovariansen γ_1 . Autokorrelationerna för a_t^2 avtar alltså exponentiellt med γ_1 , jämför detta med GARCH(1, 1)-modellen där autokorrelationerna avtar exponentiellt med $\alpha_1 + \beta_1$.

5.1.2 Icke-linjära volatilitetsmodeller

I detta avsnitt behandlas några icke-linjära utvecklingar av ARCH-modeller. Dessa modeller är icke-linjära i den meningen att de har förmåga att skilja på positiva och negativa chocker. Denna egenskap är tilltalande eftersom vi t.ex. vet från avsnitt 1.3 att en period av hög volatilitet ofta inleds med en stor negativ chock.

EGARCH-modellen

Det första försöket att utveckla GARCH-modellen så att den kan fånga upp asymmetrier gjordes av Nelson (1991) med sina sk. EGARCH (Exponential GARCH)-modeller. För att åstadkomma detta införs den viktade chocken

$$g(\varepsilon_t) = \varphi \varepsilon_t + \psi [|\varepsilon_t| - E(|\varepsilon_t|)] \quad (5.19)$$

där φ och ψ är konstanter. Både ε_t och $|\varepsilon_t| - E(|\varepsilon_t|)$ är oberoende sekvenser med väntevärde 0, därför är $E[g(\varepsilon_t)] = 0$. Asymmetrin av $g(\varepsilon_t)$ ses om (5.19) skrivs om som

$$g(\varepsilon_t) = \begin{cases} (\varphi + \psi)\varepsilon_t - \psi E(|\varepsilon_t|) & \text{om } \varepsilon_t \geq 0 \\ (\varphi - \psi)\varepsilon_t - \psi E(|\varepsilon_t|) & \text{om } \varepsilon_t < 0 \end{cases} \quad (5.20)$$

Med denna viktfunktion för chockerna skrivs EGARCH(m, s)-modellen

$$a_t = \sigma_t \varepsilon_t, \quad \ln(\sigma_t^2) = \alpha_0 + \frac{1 + \beta_1 B + \dots + \beta_s B^s}{1 - \alpha_1 B - \dots - \alpha_m B^m} g(\varepsilon_t) \quad (5.21)$$

Vi ser i (5.21) att det finns två tydliga skillnader mellan GARCH -och EGARCH-modellen, där parametrarna $\{\alpha\}$ och $\{\beta\}$ har samma funktion som i GARCH-modellen (5.7). Dels används i EGARCH-modellen den logaritmerade betingade variansen för att släppa på villkoret om positiva parametrar. Dels tillåter funktionen $g(\varepsilon_t)$ oss att analysera asymmetriska egenskaper av en tidsserie, där eventuellt negativa chocker har annan dynamik än de positiva. I EGARCH-modellen ges de negativa chockerna vikten $\varphi - \psi$, medan de positiva chockerna ges vikten $\varphi + \psi$. För ytterligare detaljer om EGARCH-modellen och dess egenskaper se Nelson (1991).

GJR-GARCH modellen

Denna modell som föreslogs av Gosten, Jagannathan och Runkle (1993) är en alternativ metod för att beskriva de asymmetriska chockerna. Modellen härrör från GARCH(1,1)-modellen med tillägget att parametern för a_{t-1}^2 beror på huruvida a_{t-1} är positiv eller negativ, dvs

$$\sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 (1 - I[a_{t-1} > 0]) + \gamma_1 a_{t-1}^2 I[a_{t-1} > 0] + \beta_1 \sigma_{t-1}^2$$

där $I(\cdot)$ är en indikatorfunktion. Villkoren för att den betingade variansen ska vara positiv är $\alpha_0 > 0$, $(\alpha_1 + \gamma_1)/2 \geq 0$ och $\beta_1 > 0$. Om stationaritetsvillkoret, $(\alpha_1 + \gamma_1)/2 + \beta_1 < 1$, är uppfyllt gäller att den obetingade variansen av a_t är

$$\sigma^2 = \frac{\alpha_0}{1 - (\alpha_1 + \gamma_1)/2 - \beta_1}$$

Den optimala prediktorn av $y_{t+\ell}$, $\hat{y}_{t+\ell|t}$, ges av det betingade medelvärdet oavsett om chockerna a_t är betingat heteroskedastiska eller inte. Metoderna för att beräkna prognoser med linjära -och icke-linjära modeller kan alltså användas både under antagande om homoskedastiska och heteroskedastiska chocker. Vidare är ℓ -stegs prognosfelet, $\hat{a}_{t+\ell|t} = y_{t+\ell} - \hat{y}_{t+\ell|t}$, tidsvarierande.

STGARCH-modellen

En annan möjlighet då man modellerar för heteroskedasticitet är att liksom i kapitel 4 använda övergångsfunktioner för sina parametrar. Vi får då den sk. STGARCH (Smooth Transition GARCH)-modellen

$$\sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 [1 - F(a_{t-1})] + \gamma_1 a_{t-1}^2 F(a_{t-1}) + \beta \sigma_{t-1}^2 \quad (5.22)$$

där $F(a_{t-1})$ är övergångsfunktionen. Om man använder den logistiska funktionen

$$F(a_{t-1}) = \frac{1}{1 + e^{-\theta \varepsilon_{t-1}}} \quad (5.23)$$

som övergångsfunktion (där $\theta > 0$) fås den sk. LSTGARCH (Logistic STGARCH)-modellen. $F(a_{t-1})$ i (5.23) ökar monotont (se Figur 4.2) från 0 till 1 då a_{t-1} ökar vilket innebär att genomslagskraften av a_{t-1} på σ_t^2 förändras mjukt från α_1 till γ_1 . Då parametern θ ökar går $F(a_{t-1})$ mot en stegfunktion som antar värdet 0 för negativa a_{t-1} , och värdet 1 för positiva a_{t-1} . LSTGARCH-modellen antar alltså att asymmetrin i σ_t^2 enbart beror på om a_{t-1} är större eller mindre än noll. I praktiken är ofta $\alpha_1 > \gamma_1$ så att en stor negativ chock ökar den betingade variansen mer än en stor positiv chock.

Modellen (5.22) kan även användas med andra övergångsfunktioner. Om man istället använder den exponentiella övergångsfunktionen (se Figur 4.3)

$$F(a_{t-1}) = 1 - e^{-\theta \varepsilon_{t-1}^2} \quad (5.24)$$

fås den sk. ESTGARCH (Exponential STGARCH)-modellen. Användandet av övergångsfunktionen (5.24) gör att det i ESTGARCH-modellen antas att asymmetrin endast beror på chockernas storlek.

VSGARCH-modellen

Fornari och Mele (1996, 1997) menar att negativa chocker ökar den betingade variansen mer än positiva bara om storleken på chockerna är stora. För små chocker däremot finns en motsatt asymmetri, dvs. att små positiva chocker ökar den betingade variansen mer än en små negativa. För denna komplicerade form av asymmetri föreslår Fornari och Mele (1997) en volatilitetsskiftes GARCH (VSGARCH)-modell

$$\sigma_t^2 = \begin{cases} \alpha_{0,1} + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2 & \text{om } a_{t-1} < 0 \\ \alpha_{0,2} + \gamma_1 a_{t-1}^2 + \delta_1 \sigma_{t-1}^2 & \text{om } a_{t-1} > 0 \end{cases} \quad (5.25)$$

Den obetingade variansen av a_t i VSGARCH(1, 1)-modellen (5.25) kan visas vara

$$\sigma^2 = \frac{\alpha_{0,1} + \alpha_{0,2}}{2 - \alpha_1 - \gamma_1 - \beta_1 - \delta_1}$$

Toppigheten av (5.25) är större än toppigheten av en GARCH(1, 1)-modell. Parametrarna i VSGARCH(1, 1)-modellen sätts så att $\alpha_1 > \gamma_1$ och $\alpha_{0,1} + \beta_1\sigma^2 < \alpha_{0,2} + \delta_1\sigma^2$, på så sätt får små positiva chocker större effekt på σ_t^2 än små negativa chocker, med omvänt förhållande för stora chocker. Anderson, Nam och Vahid (1999) modifierade VSGARCH-modellen genom att införa en mjuk övergångsfunktion istället för stegfunktionen i (5.25), den resulterade modellen kallas ANST-GARCH (Asymmetric Nonlinear STGARCH)

$$\begin{aligned}\sigma_t^2 = & (\alpha_{0,1} + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2) [1 - F(a_{t-1})] \\ & + (\alpha_{0,2} + \gamma_1 a_{t-1}^2 + \delta_1 \sigma_{t-1}^2) F(a_{t-1})\end{aligned}$$

Där $F(a_{t-1})$ är den logistiska övergångsfunktionen (5.23).

QGARCH-modellen

Sentana (1995) introducerade ytterligare ett sätt att hantera den asymmetriska effekten av a_{t-1} på volatiliteten. Modellen kallas QGARCH (Quadratic GARCH) där QGARCH(1, 1)-modellen skrivs

$$\sigma_t^2 = \alpha_0 + \gamma_1 a_{t-1} + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \quad (5.26)$$

Den nyinförda termen $\gamma_1 a_{t-1}$ gör (i kombination med $\alpha_1 a_{t-1}^2$) att positiva och negativa chocker har olika effekt på σ_t^2 . För att se detta skrivs (5.26) om som

$$\sigma_t^2 = \alpha_0 + \left(\frac{\gamma_1}{a_{t-1}} + \alpha_1 \right) a_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \quad (5.27)$$

Vi ser i (5.27) att effekten av a_{t-1}^2 på σ_t^2 är lika med $\frac{\gamma_1}{a_{t-1}} + \alpha_1$. Om $\gamma_1 < 0$, får negativa chocker större positiv inverkan på σ_t^2 än positiva chocker av samma storlek. QGARCH-modellen har många likheter med GARCH-modellen, t.ex. har de samma obetingade varians ($\sigma^2 = \frac{\alpha_0}{1 - \alpha_1 - \beta_1}$), dessutom har de samma villkor för stationaritet och för fjärde momentets existens. Det går däremot att visa att a_t i de två modellerna inte har samma toppighet eftersom toppigheten i QGARCH-modellen även beror på asymmetriparametern γ_1 (jmf. med toppigheten i GARCH(1, 1)-modellen)

$$\kappa_a = \frac{3 [1 - (\alpha_1 + \beta_1)^2 + \gamma_1^2 (1 - \alpha_1 - \beta_1) / \alpha_0]}{1 - 2\alpha_1^2 - (\alpha_1 + \beta_1)^2} \quad (5.28)$$

Toppigheten i QGARCH-modellen är alltså en ökande funktion av γ_1 . Om

$\gamma_1 = 0$ blir förstås toppigheten lika som i GARCH-modellen. Om vi konstanthåller parametrarna α_0 , α_1 och β_1 ser vi att QGARCH-modellen har högre toppighet än GARCH-modellen ($\gamma_1^2(1 - \alpha_1 - \beta_1)/\alpha_0 > 0$).

MSW-GARCH

De tidigare nämnda modellerna i det här delkapitlet kan ses på som regim-skiftesmodeller för volatiliteten, där regimen bestäms av en observerbar variabel (a_t). Likhetererna är därför många med de modeller som hanterades i kapitel 4.1. Även här då vi har icke-linjäritet i volatilitet är det naturligt att anta att regimerna också kan bestämmas av en icke-observerbar Markovprocess s_t . Vi får då en sk. MSW-GARCH (Markov Switching GARCH)-modell

$$\sigma_t^2 = \begin{cases} \alpha_{0,1} + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2 & \text{om } s_t = 0 \\ \alpha_{0,2} + \gamma_1 a_{t-1}^2 + \delta_1 \sigma_{t-1}^2 & \text{om } s_t = 1 \end{cases}$$

där s_t är en Markovkedja med övergångsmatris (4.9). Mer information om MSW-GARCH-modellens egenskaper finns i t.ex. Dueker (1997).

5.2 Test för heteroskedasticitet

Före skattning av volatilitetsmodellerna i kapitel 5.1 bör man testa serien för betingad heteroskedasticitet. Man skiljer på test för linjär -och icke-linjär heteroskedasticitet, vilka behandlas separat i avsnitt 5.2.1 och 5.2.2.

5.2.1 Test för linjär GARCH

Det finns en uppsjö av olika test för linjär betingad heteroskedasticitet, där det mest använda är ARCH-LM testet utvecklat av Engle (1982). Som namnet antyder bygger ARCH-LM testet på Lagrange-Multiplier principen. Testet inleds med att man till tidsserien skattar den mest lämpliga regressions -eller ARMA-modellen, och låter $\{\hat{\varepsilon}_t^2\}$ vara de kvadrerade residualerna. Därefter skattas regressionsmodellen

$$\hat{\varepsilon}_t^2 = \alpha_0 + \alpha_1 \hat{\varepsilon}_{t-1}^2 + \dots + \alpha_q \hat{\varepsilon}_{t-q}^2 \quad (5.29)$$

Om det inte existerar någon ARCH -eller GARCH effekt skulle alla värden α_1 till α_q vara lika med 0. Förklaringsgraden skulle vara dålig, och

determinationskoefficienten R^2 skulle vara låg. Under nollhypotesen om homoskedasticitet ($H_0 : \alpha_1 = \dots = \alpha_q = 0$) är LM-statistikan nR^2 asymptotiskt χ^2 -fördelad med q frihetsgrader, där n är stickprovsstorleken.

Andra test för linjär heteroskedasticitet är t.ex. Whites test, Breusch-Pagans test och Goldfeld-Quandts test.

5.2.2 Test för icke-linjär GARCH

Här testas istället nollhypotesen om homoskedasticitet mot mothypotesen om asymmetriska ARCH/GARCH-effekter. Engle och Ng (1993) diskuterar ett test för att avgöra om positiva och negativa chocker har olika effekt på den betingade variansen. Detta test kallas SB(Sign Bias)-testet där S_{t-1}^- är en dummyvariabel som antar värdet 1 om $\hat{a}_{t-1}$ är negativ, och värdet 0 annars. Testet undersöker om $\hat{a}_t^2$ kan förklaras av S_{t-1}^- , $S_{t-1}^- \hat{\varepsilon}_{t-1}$ och/eller $S_{t-1}^+ \hat{\varepsilon}_{t-1}$, där $S_{t-1}^+ = 1 - S_{t-1}^-$, genom att skatta regressionen

$$\hat{a}_t^2 = \phi_0 + \phi_1 S_{t-1}^- + \phi_2 S_{t-1}^- \hat{a}_{t-1} + \phi_3 S_{t-1}^+ \hat{a}_{t-1} \quad (5.30)$$

Under nollhypotesen, $H_0 : \phi_1 = \phi_2 = \phi_3 = 0$, följer LM-statistikan nR^2 asymptotiskt en χ^2 -fördelning med 3 frihetsgrader.

Ett alternativ till SB-testet är att testa nollhypotesen mot de olika typerna av asymmetriska ARCH/GARCH-effekter diskuterade i 5.1.2. T.ex. föreslår Sentana (1995) ett test av nollhypotesen om homoskedasticitet mot mothypotesen om kvadratisk ARCH (QARCH). QARCH-modellen är förstås bara ett specialfall av QGARCH-modellen (5.26) där $\beta_1 = 0$. Den allmänna QARCH(q, q)-modellen skrivs

$$\begin{aligned} \sigma_t^2 &= \alpha_0 + \gamma_1 a_{t-1} + \gamma_2 a_{t-2} + \dots + \gamma_q a_{t-q} \\ &\quad + \alpha_1 a_{t-1}^2 + \alpha_2 a_{t-2}^2 + \dots + \alpha_q a_{t-q}^2 \end{aligned}$$

Där nollhypotesen, $H_0 : \gamma_1 = \gamma_2 = \dots = \gamma_q = \alpha_1 = \alpha_2 = \dots = \alpha_q = 0$, igen testas med nR^2 från regressionen av $\hat{\varepsilon}_t^2$ på $\hat{\varepsilon}_{t-1}, \dots, \hat{\varepsilon}_{t-q}$ och $\hat{\varepsilon}_{t-1}^2, \dots, \hat{\varepsilon}_{t-q}^2$. Här är $nR^2 \sim \chi^2(2q)$.

Hagerud (1997) föreslog ett test där mothypotesen är STARCH-modellen

$$\begin{aligned} \sigma_t^2 &= \alpha_0 + \alpha_1 a_{t-1}^2 [1 - F(a_{t-1})] + \dots + \alpha_q a_{t-q}^2 [1 - F(a_{t-q})] \\ &\quad + \gamma_1 a_{t-1}^2 F(a_{t-1}) + \dots + \gamma_q a_{t-q}^2 F(a_{t-q}) \end{aligned} \quad (5.31)$$

där $F(\cdot)$ är antingen är den logistiska funktionen (5.23) eller den exponentiella funktionen (5.24). En försvårande omständighet då vi testar nollhypotesen, $H_0 : \alpha_1 = \dots = \alpha_q = \gamma_1 = \dots = \gamma_q = 0$, i (5.31), är att parametern θ i $F(\cdot)$ är oidentifierad under H_0 . Detta identifieringsproblem är liknande det vi hade i avsnitt 4.3.2. Problemet kan dock lösas med Taylorapproximering. I fallet med LSTARCH-modell (5.22 med 5.23) resulterar denna Taylorapproximering i

$$\sigma_t^2 = \alpha_0 + \alpha_1^* \hat{\varepsilon}_{t-1}^2 + \dots + \alpha_q^* \hat{\varepsilon}_{t-q}^2 + \gamma_1^* \hat{\varepsilon}_{t-1}^3 + \dots + \gamma_q^* \hat{\varepsilon}_{t-q}^3 \quad (5.32)$$

Ett LM-test för att testa $H_0 : \alpha_1^* = \dots = \alpha_q^* = \gamma_1^* = \dots = \gamma_q^* = 0$, fås som nR^2 från regressionen

$$\hat{\varepsilon}_t^2 = \alpha_0 + \alpha_1^* \hat{\varepsilon}_{t-1}^2 + \dots + \alpha_q^* \hat{\varepsilon}_{t-q}^2 + \gamma_1^* \hat{\varepsilon}_{t-1}^3 + \dots + \gamma_q^* \hat{\varepsilon}_{t-q}^3 \quad (5.33)$$

Under nollhypotesen följer denna LM-statistika en χ^2 -fördelning med $2q$ frihetsgrader.

I fallet med en ESTARCH-modell (5.22 med 5.24) får vi en modell liknande (5.32) och (5.33), men där $\hat{\varepsilon}_{t-i}^3$ ersätts med $\hat{\varepsilon}_{t-i}^4$ där $i = 1, 2, \dots, q$.

5.3 Estimering

Den betingade variansen σ_t^2 för a_t antas i detta avsnitt följa en linjär -eller en icke-linjär GARCH modell med parametervektorn $\boldsymbol{\psi}$. T.ex. är i VSGARCH-modellen (5.25) $\boldsymbol{\psi} = (\alpha_{0,1}, \alpha_{0,2}, \alpha_1, \beta_1, \gamma_1, \delta_1)'$. Dessutom skattas förstås parametrarna för det betingade medelvärdet, $\boldsymbol{\xi}$, som eventuellt består av de autoregressiva parametrarna $\phi_1, \dots, \phi_p$. Parametervektorn att skatta betecknas $\boldsymbol{\theta} = (\boldsymbol{\xi}', \boldsymbol{\psi}')'$, vilken estimeras med ML-tekniker. Den betingade log-likelihoodfunktionen för observation t är

$$\ln L(\boldsymbol{\theta}) = \ln f(a_t/\sigma_t) - \ln \sigma_t \quad (5.34)$$

där $f(\cdot)$ är täthetsfunktionen av de oberoende och identiskt fördelade chockerna, ε_t . Om dessa t.ex. antas vara normalfördelade fås

$$\ln L(\boldsymbol{\theta}) = -\frac{1}{2} (\ln 2\pi + \ln \sigma_t + a_t^2/\sigma_t^2) \quad (5.35)$$

ML skattningen av θ , $\hat{\theta}_{ML}$, fås genom att maximera (5.35) för hela stickprovet, skattningarna ges då som lösningen till

$$\sum_{t=1}^n \frac{\partial \ln L(\theta)}{\partial \theta} = 0 \quad (5.36)$$

Om ε_t är normalfördelade och det betingade medelvärdet och variansen är rätt specificerade, är $\hat{\theta}_{ML}$ konsistent, asymptotiskt normalfördelad samt har minst varians av alla konsistenta skattningar.

Som vi vet från kap. 1 är det dock inte särskilt realistiskt att tro att ε_t är normalfördelade. Den under villkoren (5.35) och (5.36) skattade modellen skulle t.ex. inte lyckas fånga upp den höga toppigheten som förekommer i många finansiella tidsserier. Istället antas ofta att ε_t följer någon annan fördelning med tjockare svansar än normalfördelningen (t.ex. t -fördelningen). Modellen skattas på samma sätt genom att maximera log-likelihoodfunktionen för den specifika fördelningen.

5.4 Diagnostik

Av samma anledning som det är viktigt att diagnosticera modeller för det betingade medelvärdet är det att diagnosticera modeller för den betingade variansen.

Kom ihåg att $a_t = \sigma_t \varepsilon_t$ i modellerna beskrivna i avsnitt 5.1. Ett antagande vi gör är att de standardiserade residualerna $\hat{\varepsilon}_t = \hat{a}_t / \hat{\sigma}_t$ är o.i.f, dvs om modellen är korrekt specificerad kommer $\hat{\varepsilon}_t$ ha konstant varians, vara oberoende etc. För att testa detta används ett antal standardtest (F -test, Ljung-Box test etc.).

Av speciellt intresse är här att se om $\hat{\varepsilon}_t$ fortfarande visar spår av betingad heteroskedasticitet. De två mest kända testen för heteroskedasticitet i $\hat{\varepsilon}_t$ är McLeod och Li (4.10) och ARCH-LM testet (5.29). På senare tid har varianter på dessa två test föreslagits. T.ex. förelögs av Lundbergh och Teräsvirta (1998) ett ARCH-LM test för ARCH(m)-effekter i $\hat{\varepsilon}_t$. Teststatistikan är nR^2 , där R^2 är determinationskoefficienten från regressionen

$$\hat{\varepsilon}_t^2 = \phi_0 + \phi_1 \hat{\varepsilon}_{t-1} + \dots + \phi_m \hat{\varepsilon}_{t-m}^2 + \lambda' \mathbf{x}_t + u_t \quad (5.37)$$

där $\mathbf{x}_t$ är en vektor bestående av partiella derivator av σ_t^2 med avseende på GARCH-modellens parametrar. Under nollhypotesen, $H_0 : \phi_1 = \dots = \phi_m$, i

(5.37) följer nR^2 asymptotiskt en χ^2 -fördelning med m frihetsgrader.

Ett liknande test föreslogs av Bollerslev (1986) för att testa för högre ordningars GARCH-modeller. För att testa specificeringen $\text{GARCH}(p, q)$ mot $\text{GARCH}(p+r, q)$ eller $\text{GARCH}(p, q+s)$ används på nytt ett LM-test, nR^2 , där R^2 kommer från

$$\hat{\varepsilon}_t^2 = \phi_0 + \phi_1 \hat{a}_{t-q-1}^2 + \dots + \phi_r \hat{a}_{t-q-r}^2 + \boldsymbol{\lambda}' \hat{\mathbf{x}}_t + u_t$$

eller

$$\hat{\varepsilon}_t^2 = \phi_0 + \phi_1 \hat{\sigma}_{t-p-1}^2 + \dots + \phi_s \hat{\sigma}_{t-p-s}^2 + \boldsymbol{\lambda}' \hat{\mathbf{x}}_t + u_t$$

Det är naturligt att överväga icke-linjära volatilitetsmodeller först då ovanstående diagnostiska test visar på asymmetri i $\hat{\varepsilon}_t$. SB-testet (5.30) kan här användas för att testa nollhypotesen om linjär GARCH mot mothypotesen om icke-linjära alternativ. Skillnaden här är att vi gör en regression på $\hat{\varepsilon}_t^2$ där de partiella derivatorna, $\hat{\mathbf{x}}_t = \partial \hat{\sigma}_t^2 / \partial \boldsymbol{\theta}$, inkluderas som regressorer

$$\hat{\varepsilon}_t^2 = \phi_0 + \phi_1 S_{t-1}^- + \phi_2 S_{t-1}^- \hat{a}_{t-1} + \phi_3 S_{t-1}^+ \hat{a}_{t-1} + \boldsymbol{\lambda}' \hat{\mathbf{x}}_t + \xi_t \quad (5.38)$$

Hagerud (1997) undersöker med simuleringstekniker SB-testets förmåga att finna olika asymmetriska GARCH-effekter. Generellt har testet låg styrka, dessutom ger förkastande av nollhypotesen inte mycket information om vilken icke-linjär modell som är det korrekta alternativet. Av denna anledning utvecklade Hagerud (1997) två test för att testa $\text{GARCH}(1, 1)$ mot de specifika alternativen $\text{QGARCH}(1, 1)$ och $\text{LSTGARCH}(1, 1)$. Dessa test är båda av LM-typ, men saknar även de förmåga att avgöra den korrekta formen av icke-linjäritet. Testen utesluter inte möjligheten att förkastandet av nollhypotesen beror på någon annan form av icke-linjär GARCH-modell. Däremot kan förstås dessa test användas för att testa nollhypotesen mot generell asymmetri.

GARCH-modeller skattas ofta på långa tidsserier, vilket väcker frågan huruvida parametrarna verkligen är konstanta under hela perioden. Chu (1995) testar nollhypotesen om konstanta parametrar mot mothypotesen om att det skett exakt ett parameterskifte (structural break), medan Lundbergh och Teräsvirta (1998) testar mot alternativet om långsamt förändrade parametervärden. I det senare fallet testas om $\text{GARCH}(1, 1)$ -modellens parametrar är konstanta mot alternativet

$$\begin{aligned}\sigma_t^2 &= (\alpha_{0,1} + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2) [1 - F(t)] \\ &\quad + (\alpha_{0,2} + \gamma_1 a_{t-1}^2 + \delta_1 \sigma_{t-1}^2) F(t)\end{aligned}\quad (5.39)$$

där $F(t)$ är den logistiska funktionen (5.23). Notera att (5.39) är väldigt lik ANST-GARCH-modellen föreslagen av Anderson, Nam och Vahid (1999). Parametern θ i $F(t)$ är inte definierad under nollhypotesen, $H_0 : \alpha_{0,1} = \alpha_{0,2}$, $\alpha_1 = \gamma_1$, $\beta_1 = \delta_1$, vilket på nytt löses genom att Taylorapproximera $F(t)$

$$\sigma_t^2 = \alpha_{0,1}^* + \alpha_1^* a_{t-1}^2 + \beta_1^* \sigma_{t-1}^2 + \alpha_{0,2}^* t + \gamma_1^* a_{t-1}^2 t + \delta_1^* \sigma_{t-1}^2 t \quad (5.40)$$

Nollhypotesen, $H_0 : \alpha_{0,2}^* = \gamma_1^* = \delta_1^* = 0$, testas nu enkelt med ett LM-test.

5.5 Prognoser

Bland de linjära volatilitetsmodellerna är det naturligt att börja med att studera volatilitetsprognoserna från ARCH-modellen. Enstegsprognoserna, $E(\sigma_{t+1}^2 | \Omega_t)$, ges direkt av ARCH(m)-modellen

$$\sigma_{t+1}^2 = \alpha_0 + \alpha_1 a_t^2 + \alpha_2 a_{t-1}^2 + \dots + \alpha_m a_{t-m+1}^2$$

Volatilitetsprognosen för tidpunkten $t + 1$ är alltså en viktad summa av a_t (plus konstanten α_0). Om vi istället antar lika vikter, $\alpha_i = 1/m$, för $i = 1, 2, \dots, m$, samt att $\alpha_0 = 0$, fås prognosen för σ_{t+1}^2 som

$$\sigma_{t+1}^2 = \frac{1}{m} \sum_{s=0}^{m-1} a_{t-s}^2$$

σ_{t+1}^2 beräknas alltså i detta fall som medelvärde av de m senaste kvadrerade chockerna. Att ge varje chock samma vikt har dock vissa nackdelar, t.ex. är $\sigma_{t+1}^2 = a_t^2$ då $m = 1$. Höga värden på a_t^2 resulterar i stor prognosticerad volatilitet i all framtid, och vice versa. Chockerna har alltså i detta fall maximal persistens.

Prognoser med ARCH(1)-modellen är pedagogiskt tilltalande. Kom ihåg att

den obetingade variansen $\sigma^2 = \frac{\alpha_0}{1-\alpha_1}$, eller ekvivalent $\alpha_0 = \sigma^2(1-\alpha_1)$, vilket gör att (5.5) kan skrivas om som

$$\sigma_t^2 = \sigma^2(1-\alpha_1) + \alpha_1 a_{t-1}^2$$

För tidpunkt $t + \ell$ fås

$$\sigma_{t+\ell}^2 = \alpha_1 (\hat{a}_{t+\ell-1}^2 - \sigma^2)$$

För att beräkna dessa prognoser tas det betingade väntevärdet av $\sigma_{t+\ell}^2$ baserat på informationen Ω_t

$$\begin{aligned} \sigma_{t+\ell|t}^2 - \sigma^2 &= E(\sigma_{t+\ell}^2 | \Omega_t) - \sigma^2 = \alpha_1 (E(\sigma_{t+\ell-1}^2 \varepsilon_{t+\ell-1}^2 | \Omega_t) - \sigma^2) \\ &= \alpha_1 (\sigma_{t+\ell-1|t}^2 - \sigma^2) \end{aligned}$$

där $a_{t+\ell-1}^2 = \sigma_{t+\ell-1}^2 \varepsilon_{t+\ell-1}^2$ (eftersom $a_t = \sigma_t \varepsilon_t$), och där

$$E(a_{t+\ell-1}^2 | \Omega_t) = E(\sigma_{t+\ell-1}^2 | \Omega_t) E(\varepsilon_{t+\ell-1}^2 | \Omega_t) = E(\sigma_{t+\ell-1}^2 | \Omega_t) = \sigma_{t+\ell-1|t}^2$$

för $\ell > 2$. Genom substituering bakåt fås att

$$\sigma_{t+\ell|t}^2 - \sigma^2 = \alpha_1^{\ell-1} (\sigma_{t+1|t}^2 - \sigma^2) = \alpha_1^\ell (a_t^2 - \sigma^2)$$

där $\sigma_{t+1|t}^2 = \sigma_{t+1}^2$. Ett viktigt resultat av detta är att

$$\sigma_{t+\ell|t}^2 \rightarrow \sigma^2, \quad \text{då } \ell \rightarrow \infty$$

givet att processen är stationär ($|\alpha_1| < 1$). Resultatet av att volatilitetsprognoserna asymptotiskt går mot σ^2 gäller även för den generella ARCH(m)-modellen.

GARCH(1,1)-modellen (5.9) har som bekant obetingad varians $\sigma^2 = \frac{\alpha_0}{1-(\alpha_1+\beta_1)}$. Om man i (5.9) ersätter α_0 med $\sigma^2(1-(\alpha_1+\beta_1))$ fås

$$\sigma_t^2 - \sigma^2 = \alpha_1 (a_{t-1}^2 - \sigma^2) + \beta_1 (\sigma_{t-1}^2 - \sigma^2) \quad (5.41)$$

Enstegsprognosen av (5.41), $E(\sigma_{t+1}^2 | \Omega_t) = \sigma_{t+1}^2$, fås därför som

$$\sigma_{t+1}^2 - \sigma^2 = \alpha_1 (a_t^2 - \sigma^2) + \beta_1 (\sigma_t^2 - \sigma^2) \quad (5.42)$$

Prognoser för $\sigma_{t+\ell}^2$ får genom att först skriva om (5.42) som

$$\begin{aligned} \sigma_{t+\ell}^2 - \sigma^2 &= \alpha_1 (a_{t+\ell-1}^2 - \sigma^2) + \beta_1 (\sigma_{t+\ell-1}^2 - \sigma^2) \\ &= \alpha_1 (\sigma_{t+\ell-1}^2 \varepsilon_{t+\ell-1}^2 - \sigma^2) + \beta_1 (\sigma_{t+\ell-1}^2 - \sigma^2) \end{aligned} \quad (5.43)$$

därefter beräknas det förväntade väntevärdet

$$\begin{aligned} \sigma_{t+\ell|t}^2 - \sigma^2 &= E(\sigma_{t+\ell}^2 | \Omega_t) - \sigma^2 = \alpha_1 (E(\sigma_{t+\ell-1}^2 \varepsilon_{t+\ell-1}^2 | \Omega_t) - \sigma^2) \\ &\quad + \beta_1 (E(\sigma_{t+\ell-1}^2 | \Omega_t) - \sigma^2) \\ &= (\alpha_1 + \beta_1) (E(\sigma_{t+\ell-1}^2 | \Omega_t) - \sigma^2) \end{aligned} \quad (5.44)$$

Rekursiv utveckling av (5.44) ger

$$\begin{aligned} \sigma_{t+\ell|t}^2 - \sigma^2 &= (\alpha_1 + \beta_1)^{\ell-1} (\sigma_{t+1}^2 - \sigma^2) \\ &= (\alpha_1 + \beta_1)^{\ell-1} [\alpha_1 (a_t^2 - \sigma^2) + \beta_1 (\sigma_t^2 - \sigma^2)] \end{aligned} \quad (5.45)$$

I likhet med ARCH-modeller gäller att $\sigma_{t+\ell|t}^2$ går mot σ^2 då $\ell \rightarrow \infty$ (givet att $\alpha_1 + \beta_1 < 1$).

Prognoser för den betingade variansen i icke-linjära GARCH-modeller kan även de beräknas analytiskt. Låt oss t.ex. se på GJR-GARCH(1, 1)-modellen

$$\sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 (1 - I[a_{t-1} > 0]) + \gamma_1 a_{t-1}^2 I[a_{t-1} > 0] + \beta_1 \sigma_{t-1}^2$$

där det antas att ε_t är symmetrisk runt 0. Prognosen av σ_{t+1}^2 ges direkt av modellen som

$$\sigma_{t+1}^2 = \alpha_0 + \alpha_1 a_t^2 (1 - I[a_t > 0]) + \gamma_1 a_t^2 I[a_t > 0] + \beta_1 \sigma_t^2$$

och

$$\sigma_{t+2|t}^2 = \alpha_0 + \alpha_1 E(a_{t+1}^2 (1 - I[a_{t+1} > 0])) + \gamma_1 E(a_{t+1}^2 I[a_{t+1} > 0]) + \beta_1 E(\sigma_{t+1}^2)$$

Det går att förenkla uttrycket för $\sigma_{t+2|t}^2$ genom att ε_t är symmetrisk runt 0. Då är

$$E(a_{t+1}^2 I[a_{t+1} > 0]) = E(a_{t+1}^2) E(I[a_{t+1} > 0]) = \sigma_{t+1}^2 \times \frac{1}{2}$$

eftersom $E(I[a_{t+1} > 0]) = P(a_{t+1} > 0) = 1/2$. Dvs,

$$\sigma_{t+2|t}^2 = \alpha_0 = ((\alpha_1 + \gamma_1)/2 + \beta_1) \sigma_{t+1}^2$$

Generellt fås en prognos för den betingade variansen vid tidpunkt $t + \ell$ som

$$\begin{aligned} \sigma_{t+\ell|t}^2 &= \alpha_0 + \alpha_1 \sigma_{t+\ell-1|t}^2 E(1 - I[a_{t+\ell-1} > 0]) \\ &\quad + \gamma_1 \sigma_{t+\ell-1|t}^2 E(I[a_{t+\ell-1} > 0]) + \beta_1 E(\sigma_{t+\ell-1|t}^2) \\ &= \alpha_0 + ((\alpha_1 + \gamma_1)/2 + \beta_1) \sigma_{t+\ell-1|t}^2 \end{aligned}$$

vilket med substitution bakåt ger

$$\begin{aligned} \sigma_{t+\ell|t}^2 &= \alpha_0 + ((\alpha_1 + \gamma_1)/2 + \beta_1) \sigma_{t+\ell-1|t}^2 \\ &= \alpha_0 + ((\alpha_1 + \gamma_1)/2 + \beta_1) (\alpha_0 + ((\alpha_1 + \gamma_1)/2 + \beta_1) \sigma_{t+\ell-2|t}^2) \\ &\quad \vdots \\ &= \alpha_0 \sum_{i=0}^{\ell-2} ((\alpha_1 + \gamma_1)/2 + \beta_1)^i + ((\alpha_1 + \gamma_1)/2 + \beta_1)^{\ell-1} \sigma_{t+1}^2 \end{aligned}$$

Genom att utnyttja att

$$\alpha_0 \sum_{i=0}^{\ell-2} ((\alpha_1 + \gamma_1)/2 + \beta_1)^i = \alpha_0 \frac{1 - ((\alpha_1 + \gamma_1)/2 + \beta_1)^{\ell-1}}{1 - ((\alpha_1 + \gamma_1)/2 + \beta_1)}$$

för $|(\alpha_1 + \gamma_1)/2 + \beta_1| < 1$, och att $\sigma^2 = \alpha_0 [1 - ((\alpha_1 + \gamma_1)/2 + \beta_1)^{-1}]$ fås ett uttryck för $\sigma_{t+\ell|t}^2$ vilket är direkt jämförbart med den för den linjära GARCH-modellen (5.45)

$$\sigma_{t+\ell|t}^2 - \sigma^2 = ((\alpha_1 + \gamma_1)/2 + \beta_1)^{\ell-1} (\sigma_{t+1}^2 - \sigma^2)$$

I QGARCH(1,1)-modellen (5.26) påverkar inte asymmetristermen $\gamma_1 a_{t-1}$ prognoserna för den betingade variansen. Detta eftersom det betingade väntevärdet av a_{t+i} , där $i > 0$, enligt antagande är lika med 0. Prognosfelet $v_{t+\ell|t} = \sigma_{t+\ell}^2 - \sigma_{t+\ell|t}^2$ blir

$$\begin{aligned} v_{t+\ell|t} &= \gamma_1 a_{t+\ell-1} + \alpha_1 (a_{t+\ell-1}^2 - a_{t+\ell-1|t}^2) + \beta_1 (\sigma_{t+\ell}^2 - \sigma_{t+\ell|t}^2) \\ &= \gamma_1 a_{t+\ell-1} + \alpha_1 v_{t+\ell-1} + (\alpha_1 + \beta_1) v_{t+\ell-1|t} \end{aligned}$$

vilket efter rekursiv substitution resulterar i

$$v_{t+\ell|t} = \alpha_1 \sum_{i=1}^{\ell-1} (\alpha_1 + \beta_1)^{i-1} v_{t+i-1} + \gamma_1 \sum_{i=1}^{\ell-1} (\alpha_1 + \beta_1)^{i-1} a_{t+i-1}$$

Eftersom $E(v_{t+\ell-1} | \Omega_t) = E(a_{t+\ell-1} | \Omega_t) = 0$ för alla $i = 1, 2, \dots, \ell - 1$, är prognoserna väntevärdesriktiga, dvs. $E(v_{t+\ell|t} | \Omega_t) = 0$. Däremot gäller att den betingade variansen av $v_{t+\ell|t}$, och därigenom osäkerheten om $\sigma_{t+\ell}^2$, är större än för GARCH(1,1)-modellen.

ESTGARCH-modellen som ges av (5.22) med övergångsfunktion (5.24), är den enda av de icke-linjära volatilitetsmodellerna som behandlades i avsnitt 5.1.2, där prognoserna av $\sigma_{t+\ell}^2$ inte kan beräknas analytiskt. Problemet är att $F(a_{t+1}) = 1 - e^{-\theta a_{t+1}^2}$ är korrelerad med a_{t+1}^2 , och att $E[F(a_{t+1})] \neq F[E(a_{t+1})]$. Det är därför inte möjligt att i det här fallet erhålla ett rekursivt uttryck för $\sigma_{t+\ell}^2$. Istället fås prognoserna för framtida betingad varians genom simulering.

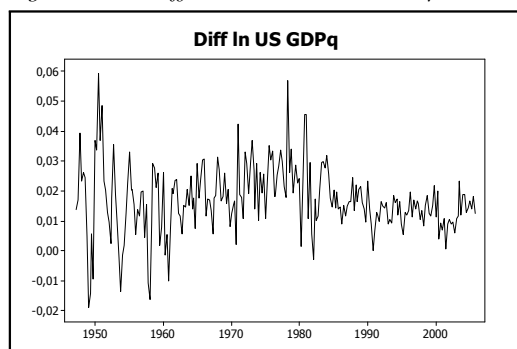
Slutligen fås prognoser av $\sigma_{t+\ell}^2$ med MSW-GARCH modeller genom att utnyttja egenskaper av Markovprocessen. Se t.ex. Dueker (1997) för mer information.

5.6 Alternativa metoder

Stockhammar och Öller (2007) föreslår ett alternativ till att modellera heteroskedasticiteten explicit genom att filtrera bort densamma och istället analysera den filtrerade homoskedastiska serien. Det finns flera fördelar med denna metod, den kanske viktigaste är att vi kan använda enklare modeller och därigenom spara på antalet parametrar. Idén är alltså i den meningen

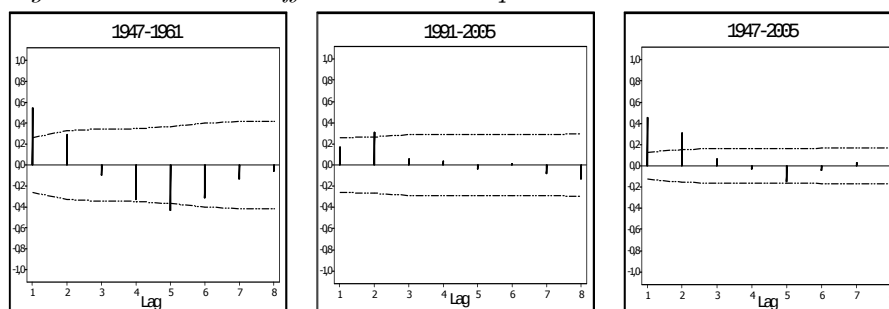
samma som vid säsongrensning. En annan viktig fördel är att man med denna metod kan få bättre parameterskattningar av skäl som framgår nedan. Låt oss studera den heteroskedastiska serien Diff. Ln. US GDP 1947-2005.

Figur 5.2: Diff. Ln. US GDP 1947-2005.



Vi ser i Figur 5.2 att serien har en avtagande volatilitet genom hela perioden. I en så lång serie är det vanligt att autokorrelationsstrukturen förändras, vilket försvårar anpassning av modeller till heteroskedastiska serier. I detta fall ges nämligen de olika autokorrelationsstrukturerna olika stor vikt vid skattningen av parametrarna, och speciellt ges observationerna i den mest volatila perioden störst vikt och vice versa. Låt oss se på autokorrelationsfunktionen av Diff. Ln. US GDP 1947-2005, samt de första och sista 15 åren (dvs. den mest och minst volatila perioden)

Figur 5.3: ACF av Diff. Ln. US GDPq



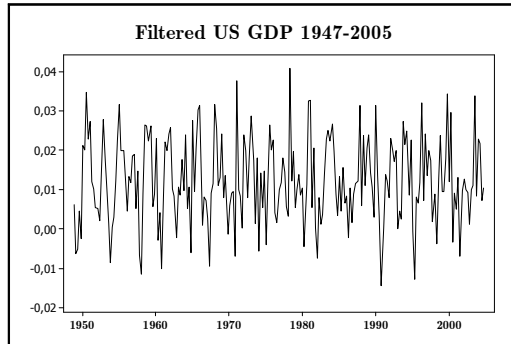
Figur 5.3 visar tydligt hur autokorrelationsstrukturen förändrats, och att ACF de första 15 åren och hela perioden är rätt lika. Den mest volatila perioden verkar alltså i stor utsträckning påverka den totala ACF. Situationen med Diff. Ln. US GDP är extra ogynnsam eftersom vi har en avtagande

volatilitet, dvs. de äldsta observationerna ges störst vikt och parameterskattningarna baseras därför allför mycket på en obsolet autokorrelationsstruktur. Optimalt vill man ge varje observation i tidsserien samma vikt, och ett sätt att åstadkomma detta är genom heteroskedasticitetsfiltrering. Ibid. föreslog följande ad hoc filter

$$y_t^* = s_y \left[\frac{y_t - \sum_{\tau=t-\eta}^{t+\eta} \frac{y_\tau}{k}}{HP^{(\lambda)} \left\{ \frac{\sum_{\tau=t-\eta}^{t+\eta} \left(y_t - \sum_{\varphi=t-\eta}^{t+\eta} \frac{y_\varphi}{k} \right)^2}{2\eta} \right\}^{1/2}} \right] + \bar{y}, \quad t = k-\eta, \dots, n-\eta$$

där y_t = Diff. Ln. GDP, $\eta = (k-1)/2$, y_t^* är den filtrerade serien, k är fönsterlängden, $\bar{y}$ och s_y är hela periodens medelvärde respektive standardavvikelse. I täljaren subtraheras ett k -termers centrerat glidande medelvärde för att extrahera eventuella lokala och globala trender. I nämnaren har vi en glidande serie med standardavvikelser utjämnad med ett lämpligt Hodrick-Prescott filter. Ibid. visade att detta filter lyckades filtrera bort heteroskedasticiteten i BNP-serierna för USA, Storbritanien och Australien. T.ex. blir Diff. Ln. US GDP efter filtrering

Figur 5.4: Diff. Ln. US GDP 1947-2005, filtrerad

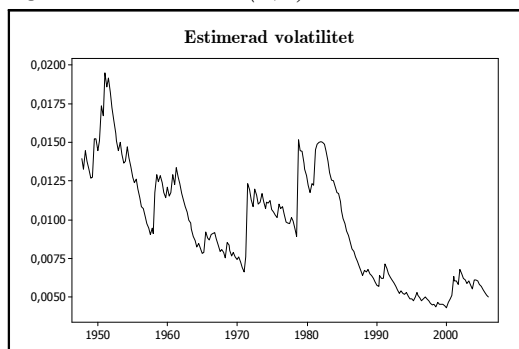


Egenskaperna hos vitt brus påverkas inte av filtret, dvs. det förorsakas inga oönskade effekter på seriens dynamiska egenskaper. Om man analyserar den homoskedastiska filtrerade serien i Figur 5.4 med ARMA-modeller upptäcktes en mycket intressant bonuseffekt. Det visade sig att man för samtliga

serier och prognoshorisonter (1,2,3 och 5 år) får signifikant bättre punktprognoser med denna metod jämfört med ARMA och ARFIMA-modeller baserade på den heteroskedastiska serien. Dessutom fås mindre standardfel för prognoserna jämfört med ARMA och ARFIMA, och ungefär samma som med GARCH-modeller. Detta kan härröra sig från filtreringen och det utjämnade viktsystem som denna medför. Efter filtreringen kan vi alltså använda enkla linjära ARMA-modeller med samma eller i många fall bättre prognosförmåga än mer komplicerade modeller.

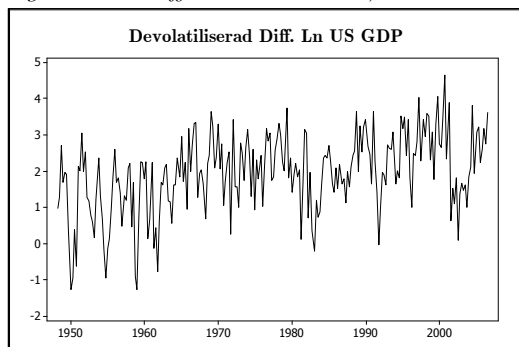
Ett annan naturlig metod att rensa en tidsserie på heteroskedasticitet används av bland andra Zhang (2006) och Engle (2002). De tillämpar en sk. devolatiliserad serie som fås genom att helt enkelt dividera den heteroskedastiska serien med den av en GARCH(1,1)-modells skattade volatilitet. I figur 5.5 ses den skattade volatiliteten för Diff. Ln. US GDP, 1947-2005

Figur 5.5: GARCH(1,1)-skattad volatilitet av Diff. Ln. US GDP



Vi ser i Figur 5.5 att volatiliteten är avtagande. Om nu den heteroskedastiska serien Diff. Ln. US GDP i Figur 5.2 divideras med den skattade volatiliteten i Figur 5.5 fås den devolatiliserade serien

Figur 5.6: Diff. Ln US GDP, devolatiliserad



Devolatiliseringen verkar rensa serien på heteroskedasticitet, men tillför en svag (icke-signifikant) positiv trend. Denna trend kan extraheras genom att beräkna 2:a differenserna av serien eller genom att som i Stockhammar och Öller (2007), subtrahera en glidande medelvärdesserie.

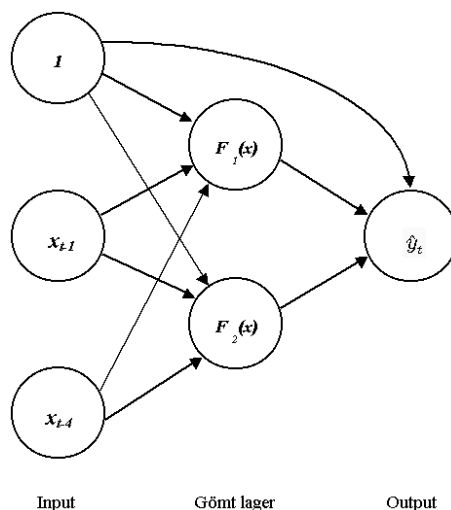
En annan ännu oprövad idé är att använda ett modellbaserat filter istället för ovanstående ad hoc filter. Man kan tänka sig att filtrering sker med tillämpning av en lämplig volatilitetsmodell (linjär eller icke-linjär). Ett problem är att den estimerade volatiliteten i Figur 5.5 inte är stationär, vilket dock ("som vanligt") löses genom skapandet av ett lämpligt antal differenser.

6 Artificiella neurala nätverk

Ett artificiellt neuralt nätverk (ANN) tar emot en signal, behandlar den, och avger resultat i form av en ny signal. ANN-modeller är en slags matematisk modell där man kan beskriva deras funktion uttömmande genom att endast ange de matematiska formlerna för signalbearbetningen. Signalerna kan vara av helt olika slag, varför ANN-modeller används inom en lång rad av olika områden som mönsterigenkänning, genetik, modellering av signaler i hjärnan etc. Dessa beräkningsintensiva modeller har på senare tid även flitigt använts på tidsseriedata. Den främsta anledningen till detta är den aldrig avtagande utvecklingen av datorkapacitet. Dessutom är ANN-modeller väldigt flexibla, skattningen av en ANN-modell till en icke-linjär tidsserie ger bättre anpassning än en linjär modell, utan att man exakt behöver specificera formen av icke-linjäritet. Några ofta nämnda nackdelar med ANN-modeller är att skattningarna är svåra att tolka, överanpassning samt att den skattade modellen inte ger någon vägledning om den korrekta formen av icke-linjäritet.

Ett ANN består dels av ett input av ett antal noder (eller neuroner), x_i , eventuellt med förbindelser som var och en har en viss vikt w . Dessa noder går sedan via ett antal "gömda skikt" till output. Styrkan på förbindelsen från den i :te noden i input till den j :te noden i det gömda skiktet har vikten w_{ij} . Dessa gömda lager består även de av en eller flera noder och förbinder alltså på ett eventuellt icke-linjärt sätt input med output. Nätverket klassificeras som antingen framåt -eller bakåtkopplat (feed-forward eller "feed-back") beroende på om det finns återkopplingar inom samma skikt. I Figur 6.1 nedan ses ett exempel på ett artificiellt nätverk.

Figur 6.1 Exempel på ett artificiellt neuralt nätverk



Figur 6.1 visar ett nätverk med tre inputs, ett gömt skikt med två noder och ett output. Det är ett framåtkopplat nätverk eftersom det inte finns några förbindelser inom samma skikt. Ibland används klassificeringen ”3-2-1 framåtkopplat nätverk” för strukturen i Figur 6.1. Inputlagret består i det här exemplet av en konstant tillsammans med x_{t-1} och x_{t-4} , vilket gör att denna struktur är naturlig att använda på kvartalsdata. Informationen rör sig från input till output via det gömda skiktet, som i Figur 6.1 består av två sk. aktiveringsfunktioner, $F_1(x)$ och $F_2(x)$. Nätverkets egenskaper bestäms alltså i hög grad av vilken aktiveringsfunktion som används. Aktiveringsfunktionen kan vara linjär eller en stegfunktion, men vanligast är att den är någon sorts sigmoidfunktion (logistisk, hyperbolisk, ...).

I det generella fallet med ett framåtkopplat nätverk och ett gömt lager kan den j :te noden definieras som

$$H_j = F_j \left(\alpha_{0j} + \sum_{i \rightarrow j} w_{ij} x_i \right) \quad (6.1)$$

där x_i är värdet på den i :te inputnoden, och där $F_j(\cdot)$ vanligen är den logistiska funktionen

$$F_j(z) = \frac{e^z}{1 + e^z}$$

α_{0j} kallas av datorvetare och i litteraturen för bias. Statistiker däremot kallar α_{0j} för dess "rätta" namn - intercept. Summeringen $i \rightarrow j$ innebär att man summerar över alla inputnoder som leder till j , w_{ij} är vikterna. Vi illustrerar detta på det 3-2-1 framåtkopplade nätverket i Figur 6.1, där den j :te noden är

$$H_j = \frac{e^{\alpha_{0j} + w_0 + w_{1j}x_{t-1} + w_{2j}x_{t-4}}}{1 + e^{\alpha_{0j} + w_0 + w_{1j}x_{t-1} + w_{2j}x_{t-4}}}, \quad \text{för } j = 1, 2. \quad (6.2)$$

där w_0 är konstanten multiplicerat med sin vikt. För outputlagret, $\hat{y}_t$, definieras denna nod som

$$\hat{y}_t = F_{\hat{y}_t} \left(\alpha_{0_{\hat{y}_t}} + \sum_{j \rightarrow \hat{y}_t} w_{j\hat{y}_t} H_j \right) \quad (6.3)$$

I många tillämpningar används samma aktiveringsfunktion för det gömda skiktet ($F_j(\cdot)$) och output ($F_{\hat{y}_t}(\cdot)$), men vid prognoser av tidsseriesdata kan det få katastrofala konsekvenser. T.ex. ger den logistiska funktionen ett tal mellan 0 och 1, och fungerar alltså bara på data som är skalad mellan 0 och 1. Av denna anledning används ofta en linjär funktion eller en indikatorfunktion för $F_{\hat{y}_t}(\cdot)$ (där $F_{\hat{y}_t} = 1$ om $z > 0$, annars är $F_{\hat{y}_t} = 0$). Om $F_{\hat{y}_t}(\cdot)$ är linjär gäller att

$$\hat{y}_t = \alpha_{0_{\hat{y}_t}} + \sum_{j=1}^k w_{j\hat{y}_t} H_j \quad (6.4)$$

där k är antalet noder i det gömda skiktet. (6.4) kan alltså ekvivalent skrivas som

$$\hat{y}_t = \alpha_{0_{\hat{y}_t}} + w_0 + w_{1\hat{y}_t} H_1 + w_{2\hat{y}_t} H_2$$

Om $F_{\hat{y}_t}(\cdot)$ är en indikatorfunktion fås istället

$$\hat{y}_t = \begin{cases} 1 & \text{om } \alpha_{0_{\hat{y}_t}} + w_0 + w_{1\hat{y}_t} H_1 + w_{2\hat{y}_t} H_2 > 0 \\ 0 & \text{om } \alpha_{0_{\hat{y}_t}} + w_0 + w_{1\hat{y}_t} H_1 + w_{2\hat{y}_t} H_2 \leq 0 \end{cases}$$

Om vi kombinerar de två skikten, kan output av ett framåtkopplat nätverk

skrivas som

$$\hat{y}_t = F_{\hat{y}_t} \left[\alpha_{0_{\hat{y}_t}} + \sum_{j \rightarrow \hat{y}_t} w_{j\hat{y}_t} F_j \left(\alpha_{0_j} + \sum_{i \rightarrow j} w_{ij} x_i \right) \right] \quad (6.5)$$

Om man också tillåter direkta förbindelser mellan input och output (som i Figur 6.1) kan nätverket skrivas

$$\hat{y}_t = F_{\hat{y}_t} \left[\alpha_{0_{\hat{y}_t}} + \sum_{i \rightarrow \hat{y}_t} \alpha_{i\hat{y}_t} x_i + \sum_{j \rightarrow \hat{y}_t} w_{j\hat{y}_t} F_j \left(\alpha_{0_j} + \sum_{i \rightarrow j} w_{ij} x_i \right) \right] \quad (6.6)$$

Nätverk som (6.6) där vissa förbindelser går direkt från input till output, och på så sätt ”hoppas över” det gömda skiktet, kallas ”skip-layer” nätverk. ANN-modeller är semiparametriska av skäl som syns i ekvation (6.5) och (6.6). Den funktionella formen är känd, men antalet noder och dess intercept är okänt.

Tillämpning av neurala nätverk sker i regel i två steg, där man i det första steget ”tränar” nätverket. Med detta menas att försöka bestämma antalet noder och skatta intercept och vikt. I steg 2 utförs inferens, speciellt prognoser. I träningsfasen delas data in i två icke-överlappande delar, där den första delen används till att skatta parametrarna för ett givet nätverk. Därefter används den andra delen av data för prognoser och beräkandet av prognosförmåga. Efter att ha genomfört dessa två steg på ett antal nätverk väljs sedan det nätverk som producerar bäst prognoser enligt ett på förhand valt kriterium. Olika algoritmer har föreslagits för att skatta ANN-modeller, men på grund av de många parametrarna är det vanligt att det tar flera tusen iterationer innan konvergens. Se t.ex. på det relativt enkla nätverket i Figur 6.1 där vi har nio förbindelser, och därför nio vikter att skatta. Det stora antalet parametrar gör att det föreligger risk att ”överträna” data och felaktigt erhålla en god anpassning inom stickprovet, men inte nödvändigtvis till bättre prognoser. Detta motiverar förstås användandet av AIC eller BIC där vi bestraffar tillägget av extra parametrar.

Det är faktiskt svårt att göra en rättvis jämförelse mellan ANN-modeller och andra tidsseriemodeller vad gäller prognosförmåga. Detta eftersom statistiker i regel inte har speciellt god kunskap om ANN-modeller, medan datorvetare å andra sidan har liten kunskap om statistiska alternativ. Weigend och Gershenfeld (1994) gjorde en empirisk studie där sex serier analyserades med stor mängd olika modeller (inkl. ANN-modeller). Serierna var alla

väldigt långa (drygt 30 000 obs.) och åtminstone fem av dem var klart icke-linjära. Det intressanta resultatet av denna studie var att ANN-modellerna fungerade väldigt bra för vissa serier, men för andra serier visade sig ANN-prognoserna vara bland de sämsta. Andra studier på tidsserier av varierande längd tyder på att ANN-modeller fungerar (relativt andra tidsseriemodeller) sämre på korta serier av linjär karaktär, och bättre på långa icke-linjära tidsserier.

7 Sammanfattning

I det här kompendiet har vi sett på ett antal icke-linjära modeller ämnade att beskriva främst finansiella tidsserier. Anledningarna till varför dessa tidsserier ofta måste modelleras med icke-linjära modeller är flera. I kapitel 1 såg vi t.ex. att finansiella tidsserier ofta är heteroskedastiska samt att stora förändringar, och speciellt stora negativa förändringar, är vanligare än förväntat (under antagande om normalfördelning). Som tidigare nämnts är forskningen inom icke-linjära tidsseriemodeller exponentiellt ökande (också det en icke-linjär funktion), vilket innebär att det finns en oerhörd mängd artiklar och annan litteratur skrivna på området. Modellerna i kapitel 3-6 i denna skrift är därför bara ett urval av de mest använda och beprövade modellerna.

I kapitel 3 såg vi på några icke-linjära utvecklingar av ARMA-processer, där den bilinjära modellen har den tilltalande egenskapen att positiva chocker är mer persistenta än negativa chocker. En simuleringsstudie visade dessutom att den bilinjära modellen har negativ skevhet och toppighet större än tre, vilket också överensstämmer med många finansiella tidsserier.

I kapitel 4 studerade vi olika regimskiftesmodeller för tidsserier som är icke-linjära i medelvärde. I TAR och SETAR-modellerna sker övergången i ett steg, eftersom man använder indikatorfunktionen som övergångsfunktion. Ett intuitivt mer tilltalande alternativ är att övergången från en regim till en annan sker mjukt, t.ex. med en logistisk eller en exponentiell funktion, vilket är fallet i LSTAR och ESTAR-modellerna. Ytterligare ett alternativ är att låta regimerna bestämmas av en icke-observerbar Markovprocess. I dessa MSW-modeller beror persistensen både på de autoregressiva parametrarna och på övergångssannolikheterna i övergångsmatrisen. Kapitlet avslutades med hur dessa regimskiftesmodeller testas, estimeras, diagnosticeras och prognosticeras.

I kapitel 5 drogs en skiljelinje mellan linjära -och icke-linjära volatilitetsmodeller. Skillnaden mellan dessa båda typer av modeller är att de linjära

volatilitetsmodellerna, till skillnad från de icke-linjära, inte har förmåga att skilja på positiva och negativa chocker. De icke-linjära volatilitetsmodellerna är av denna anledning mer tilltalande eftersom de har förmåga att beskriva de ofta förekommande asymmetrierna i finansiella tidsserier (t.ex. att stora negativa chocker är vanligare än stora positiva). Även detta kapitel avslutades med hur dessa modeller testas, estimeras, diagnosticeras och prognosticeras.

Denna skrift avslutades med kapitel 6 där de i högsta grad forskningsaktuella ANN-modellerna redovisades. Den stora fördelen med dessa modeller är att vi inte behöver specificera formen av icke-linjäritet, vilket samtidigt är en nackdel eftersom parameterskattningarna blir svåra att tolka. I simuleringstudier verkar dessa flexibla modeller ha bra prognosförmåga (gentemot andra modeller) på långa icke-linjära tidsserier. Däremot verkar ANN-modeller fungera dåligt på korta linjära tidsserier.

I ovanstående urval har vissa intressanta modeller ofrånkomligen sorterats bort och inte omnämnts. T.ex. har inget nämnts om de många modeller som tar säsongkomponenten i beaktande. Det finns en stor mängd litteratur som dokumenterar veckodags-, månads- och/eller kvartalseffekter. Dessa modeller skiljer sig dock inte mycket från de modeller som redovisats i denna skrift, och de torde vara enkla att förstå och applicera med kännedom av det material som ovan presenterats. Dessutom har inget nämnts om data som insamlats med en frekvens högre än en dag. Det finns modeller, t.ex. de sk. realiserade volatilitetsmodellerna (RV), som bygger på intradagsdata, där man i teorin kan få bättre parameterskattningar genom att utnyttja den extra information som intradagsdata medför. Däremot behandlas filtrering av heteroskedasticitet.

Det finns fortfarande stor otrampad terräng på detta intressanta område, så inget tyder på att forskningen kommer att avta den närmaste tiden. Bland annat är teorin bakom multivariata icke-linjära modeller relativt outforskad, dessutom tror man att användningen av högfrekvent data kommer öka med den förbättrade datorkapaciteten. Min förhoppning och tro är att denna text ger en solid grund att stå på inför dessa många förväntade innovationer.

Referenser

- Anderson, H. M., Nam, K. och Vahid, F. (1999) Asymmetric nonlinear smooth transition GARCH models. *Nonlinear time series analysis of economic and financial data*, Boston: Kluwer, 191-207.
- Bacon, D. W. och Watts, D. G. (1971) Estimating the transition between two intersecting straight lines. *Biometrika*, 58, 525-534.
- Baillie, R. T., Bollerslev, T och Mikkelsen, H.-O. (1996) Fractionally integrated generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics*, 74, 3-30.
- Bollerslev, T. (1986) Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics*, 31, 307-327.
- Brock, W. Dechert, D. Scheinkman, J. och Le Baron, B. (1996) A test for independence based upon the correlation dimension. *Econometric Reviews*, 15, 197-235.
- Chan, K. S. Petrucelli, J. D. Tong, H. och Woolford, S. W. (1985) A multiple threshold AR(1) model. *Journal of Applied Probability*, 22, 267-279.
- Chan, K. S. och Tong, H. (1986) On estimating thresholds in autoregressive models. *Journal of Time Series Analysis*, 7, 179-190.
- Chan, K. S. (1993) Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model. *Annals of Statistics*, 21, 520-533.
- Chu, C.-S.J. (1995) Detecting parameter shifts in GARCH models. *Econometric Reviews*, 14, 241-266.
- Ding, Z. och Granger, C. W. J. (1996) Modeling volatility persistence of speculative returns: a new approach. *Journal of Econometrics*, 73, 185-215.
- Dueker, M. J. (1997) Markov switching in GARCH processes and mean-reverting stockmarket volatility. *Journal of Business and Economic Statistics*, 15, 26-34.
- Eitrheim, Ø. och Teräsvirta, T. (1996) Testing the adequacy of smooth transition autoregressive models. *Journal of Econometrics*, 74, 59-76.
- Engle, R. F. (1982) Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50, 987-1007.

- Engle, R. F. och Bollerslev, T. (1986) Modelling the persistence of conditional variances. *Econometric Reviews*, 5, 1-50.
- Engle, R. F. Lilien, D. M. och Robins, R. P. (1987) Estimating time varying risks premia in the term structure: the GARCH-M model. *Econometrica*, 55, 391-407.
- Engle, R. F. och Ng, V. K. (1993) Measuring and testing the impact of news on volatility. *Journal of Finance*, 48, 1749-1778.
- Engle, R. F. (2002). Dynamic conditional correlation - a simple class of multivariate GARCH models. *Journal of Business and Economic Statistics*, 17, 339-350.
- Fornari, F. och Mele, A. (1996) Modeling the changing asymmetry of conditional variances. *Economics Letters*, 50, 197-203.
- Fornari, F. och Mele, A. (1997) Sign- and volatility -switching ARCH models: theory and applications to international stock markets. *Journal of Applied Econometrics*, 12, 49-65.
- Franses, P. H. och van Dijk, D. (2000) Non-linear time series models in empirical finance. Cambridge: Cambridge University Press.
- Glosten, L. R. Jagannathan, R. och Runkle, D. E. (1993) On the relation between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance*, 48, 1779-1801.
- Granger, C. W. J. och Teräsvirta, T. (1993) Modelling nonlinear economic relationships. Oxford: Oxford University Press.
- Hagerud, G. E. (1997) A new non-linear GARCH model. Ph.D. thesis, IFE, Stockholm School of Economics.
- Hamilton, J. D. (1989) A new approach to the economic analysis of nonstationary time series subject to changes in regime. *Econometrica*, 57, 357-384.
- Hamilton, J. D. (1990) Analysis of time series subject to changes in regime. *Journal of Econometrics*, 45, 39-70.
- Hamilton, J. D. (1996) Specification testing in Markov-switching time series models. *Journal of Econometrics*, 70, 127-157.
- Hansen, B. E. (1992) The likelihood ratio test under nonstandard assumptions: testing the Markov switching model of GNP. *Journal of Applied Econometrics*, 7, S61-S82; erratum (1996) 11, 195-198.

- Le Baron, B. (1992) Some relationships between volatility and correlations in stock market returns. *Journal of business*, 65, 199-219.
- Lundbergh, S. och Teräsvirta, T. (1998) Evaluating GARCH models. *Working papers in Economics and Finance* 292. Stockholm School of Economics.
- Luukkonen, R. Saikkonen, P. och Teräsvirta, T. (1988) Testing linearity against smooth transition autoregressive models. *Biometrika*, 75, 491-499.
- Mandelbrot, B. (1963) The variation of certain speculative prices. *Journal of Business*, 36, 394-419.
- Mc Culloch, R. E. och Tsay, R. S. (1993) Bayesian inference and prediction for mean and variance shifts in autoregressive time series. *Journal of the American Statistical Association*, 88, 968-978.
- McLeod, A. och Li, W. (1983) Diagnostics checking ARMA time series models using squared residual correlations. *Journal of Time Series Analysis*, 4, 269-273.
- Melino, A. och Turnbull, S. M. (1990) Pricing foreign currency options with stochastic volatility. *Journal of Econometrics*, 45, 239-265.
- Michael, P. Nobay, R. och Peel, D. (1997) Transaction costs and nonlinear adjustment in real exchange rates: an empirical investigation. *Journal of Political Economy*, 105, 862-879.
- Nelson, D. B. (1990) Stationarity and persistence in the GARCH(1, 1) model. *Econometric Theory*, 6, 318-334.
- Nelson, D. B. (1991) Inequality constraints in the univariate GARCH model. *Journal of Business and Economic Statistics*, 10, 229-235.
- Potter, S. (1995) A nonlinear approach to US GNP. *Journal of Applied Econometrics*, 10, 109-125.
- Ramsey, J. B. (1969) Tests for specification errors in classical linear least-squares regression analysis. *Journal of the Royal Statistical Society B*, 31, 350-371.
- Sentana, E. (1995) Quadratic ARCH models. *Review of Economic Studies*, 62, 639-661.
- Stockhammar, P. och Öller, L-E. (2007) A heteroscedasticity removing filter. *Research Report 2007:1*, Department of Statistics, Stockholm University.
- Taylor, S. J. (1986) Modelling financial time series, New York: John Wiley.

Tong, H. (1978) On a threshold model. Pattern Recognition and Signal processing, Amsterdam: Sijthoff & Noordhoff, 101-141.

Tong, H. (1990) Non-linear time series: A dynamical systems approach, Oxford: Oxford University Press.

Weigend, A. S. och Gershenfeld, N. A. (1994) Time series prediction. Proc. vol. XV, Santa Fe Institute in the Sciences of Complexity. Reading: Addison-Wesley.

Zhang, Z. (2006) A new class of tail-dependent time-series models and its applications in financial time series. Econometric Analysis of Financial and Economic Time Series/Part B, Advances in Econometrics, Vol 20, 317-352.