Elpriserna på den nordiska elbörsen

Prognosmodellering med hjälp av ARIMA-modeller

André Gustavsson

Student
Vt 2009
Examensarbete, 15 hp
Statistik C 30 hp
Handledare: Anders Lundquist och assistans från Peter Lindberg (Skellefteå kraft AB)

Title

Electricity prices on the Nordic power exchange

- Forecasting with the help of ARIMA-models

Abstract

The goal with this paper was to make future predictions regarding electricity spot prices acquired from the Nordic power exchange. Through autoregressive integrated moving average models I have been able to reveal the dependencies in the underlying data and later make forecasts. The models were created and based on three different types of spot prices with dummy variables to be able to analyze if there exists price fluctuations.

The finding suggests that these models can be one possible way to make future predictions when the forecasting made has proven to be fairly good when compared to real new data. Also have variation in the prices been substantiated even if some parameters have proven to be non-significant.

Making predictions through this kind of modelling is one way to do forecasting and this paper should be seen as an input in how to apply these models in real life.

Keywords: forecasting, electricity prices, ARIMA, Nordic power exchange

1. INLEDNING **3**

1.1 Bakgrund 3
1.2 Problemformulering 4
1.3 Syfte och frågeställningar 5
1.4 Avgränsningar 6
1.5 Disposition 6

2. TEORETISKT RAMVERK **7**

2.1 Tidsserier 7
2.2 ARIMA 7
2.2.1 Stationaritet *8*
2.2.2 Purely random process (Vitt Brus) *11*
2.2.3 Autoregressiv modell (AR) *12*
2.2.4 Glidande medelvärdes modell (MA) *12*
2.2.5 Sammansatt ARIMA och SARIMA *13*
2.2.6 Autokorrelations- och Partiella autokorrelations-funktionen (ACF & PACF) *14*
2.2.7 Informationskriterier *16*
2.3 Prognostisering 18

3. METOD **21**

3.1 Box-Jenkins modellbyggnadsmetodologi 21
3.2 Val av ansats 22
3.3 Urval 22
3.4 Datainsamlingen 23
3.5 Metodkritik 23
3.5.1 Validitet *23*
3.5.2 Reliabilitet *23*
3.5.3 Källkritik *23*

4. DATA **24**

4.1 Datamaterialet 24
4.1.1 Timprismodellen *24*
4.1.2 Dagsmedelprismodellen *25*
4.1.3 Månadsmedelprismodellen *25*

5. RESULTAT **26**

5.1 Timpris 26
5.2 Dagsmedelpris 31
5.3 Månadsmedelpris 37

6. DISKUSSION OCH SLUTSATSER **43**

REFERENSER **46**

BILAGOR **49**

Bilaga 1. Modellbyggnadsprocessen för timpris 49
Bilaga 2. Modellbyggnadsprocessen för dagsmedelpris 56
Bilaga 3. Modellbyggnadsprocessen för månadsmedelpriser 62

1. Inledning

Nedan följer lite bakgrundsfakta kring problemområdet och en beskrivning av uppgiften. Kapitlet avslutas med uppsatsens syfte och dess avgränsningar samt en redogörelse av upplägget.

1.1 Bakgrund

Under de senaste tjugo åren har elmarknader världen runt avreglerats och konkurrensutsatts i hopp om att kunna effektiviseras. I norden började man redan under 1960-talet inleda ett samarbete i syfte att utbyta kraft mellan varandra. Detta ledde till en omfattande utbyggnad av överföringsförbindelser mellan länderna som senare under 1990-talet möjliggjorde en utveckling mot en genensam nordisk elmarknad med den nordiska elbörsen Nord Pool som sin grund. (Heden, 2007)

Nord Pool grundades 1996 när den svenska elmarknaden avreglerades och integrerades med den tidigare avreglerade norska elmarknaden. Nord Pool var världens första multinationella marknadsplats för krafthandel med idag över 400 stycken deltagare från mer än 20 länder [1]. På Nord Pool ges möjligheten för elproducenter och elkonsumenter att handla med fysiska elkontrakt samt att prissäkra framtida säljvolymer och inköpsbehov. De aktörer som deltar inkluderar bland annat elproducenter, elleverantörer, nätbolag, finansiella institutioner, större/mindre industrier med mera. (Byström, 2000)

Marknaderna inom Nord Pool består av en spotmarknad, Elspot där handeln med fysiska elkontrakt sker samt en marknad för finansiella derivat, Eltermin. På den fysiska marknaden kan deltagarna även justera balanserna i sina elkontrakt upp till en timma före leverans i

[1] Se www.nordpool.com, hämtat 2009-06-16

korttidsmarknaden Elbas. Andra avdelningar som är aktiva inom Nord Pool är clearinginstitut och konsultavdelningar plus att det även sker handel med utsläppsrätter och kolkrediter. [2]

Elspot är, till skillnad från en vanlig börsmarknad där produkter säljs för omedelbar leverans, en ”dagen före”– marknad där handel i form av timvisa elkontrakt köps eller säljs för leverans nästkommande 24 timmar. Kontrakten köps och säljs i tre former av budgivning, nämligen i timvisa, i block och flexibla timbud. (Hjalmarsson, 2000)

Vid klockan 12 varje dag, då alla elektroniska köp- och säljbud kommit in, summeras alla bud ihop till en utbuds- och efterfrågekurva för varje leveranstimma kommande dygn. Det är en slags simultan prissättning där jämviktspriset bestäms av skärningspunkten mellan utbudet och efterfrågan. Priset i jämvikten kallas spotpris eller systempris vilket representerar alla deltagare i den nordiska regionen och används senare i de flesta fall även som referenspris på den finansiella derivatmarknaden. Priset definieras i MWh/h och säljs för EUR/MWh. (Landsorganisationen i Sverige, 2007)

Den fysiska handeln på Nord Pool står för över 60 % av den totala elkonsumtionens värde i den nordiska regionen och innefattar länder som Sverige, Norge, Finland, Danmark samt KONTEK-området i Tyskland. (Heden, 2007)

1.2 Problemformulering

På grund av svårigheten med att lagra kraft kan ansenliga samhällsekonomiska kostnader uppstå när man som till exempel elproducent inte vet exakt hur mycket elektricitet som kommer att efterfrågas och kanske därför tillverkar en felaktig mängd, det vill säga att det uppstår en utbudsbrist eller ett utbudsöverskott.

För deltagare i en sådan här fungerande elmarknad är det därför viktigt att kunna förutse ungefär hur mycket elektricitet som kommer att efterfrågas eller utbjudas och till vilket pris. En insikt i framtida spotpriser kan därför vara av stor hjälp vid planeringsprocessen.

[2] Se www.nordpool.com, hämtat 2009-06-16

För att avgöra vilket spotpris som kommer att råda kan man använda sig av prognoser. Baudin (1996) definierar syftet med prognoser som att nedbringa osäkerheten om det framtida skeendet och därmed skapa underlag för rationella beslut. Genom förenklade bilder av verkligheten i form av modeller kan man prognostisera framtida värden.

1.3 Syfte och frågeställningar

Utifrån tidigare data på elspotpriser som erhållits från Nord Pool kommer denna undersökning att konstruera prognosmodeller i form av så kallade ARIMA-modeller där beroendestrukturen i datamaterialet kommer utredas för att sedan kunna förutspå framtida värden. Elspotpriserna kommer hädanefter hänvisas till som elpriser.

När det gäller priser på elektricitet så kan det tänkas finnas en hel del fluktuationer. Många av dessa beror på förändringar i efterfrågan på kraft. Exempelvis sägs elpriserna vara lägre under vår och sommar då efterfrågan på elektricitet sjunker på grund av den ökande värmen. Inom ett dygn sägs elpriserna vara högre under morgonen och eftermiddagen på grund av alla företag som är aktiva under dagen vilket bör öka efterfrågan på elektricitet. När man ser på variationer under veckan så tros det existera en prissänkning på elektricitet under helgen då företag ej är lika aktiva och därmed inte efterfrågar lika mycket elektricitet. [3]

Utöver de nämnda prognosmodellerna kommer det även försöka utläsas de ovan nämnda teoretiskt tänkbara variationerna. Frågeställningarna uppsatsen utgår ifrån är således:

- Vad kan man som deltagare i Nord Pool vänta sig för elpriser?
- Finns det belägg för de teoretiska elprisvariationerna?

[3] Se www.nordpool.com, hämtat 2009-06-16

1.4 Avgränsningar

När man arbetar med prognoser finns det en mängd olika modeller man kan använda sig av som till exempel trendmodeller eller säsongsmodeller. Denna uppsats kommer att bestå av ARIMA-modeller.

Det kommer att skapas tre olika modeller baserat på timpriser, dagsmedelpriser samt månadsmedelpriser. Detta görs på grund av att modeller av ARIMA-typ ger prognoser med god precision på kort och medellång sikt (Baudin, 1996). På grund av detta kan prognoser göras på längre sikt genom att använda priser som sträcker sig över längre tid. I dessa tre modeller kommer det att kontrolleras de ovan nämnda säsongsvariationer.

1.5 Disposition

I Kapitel 2 presenteras de teorier som kommer att användas när datamaterialet ska analyseras och för att skapa prognosmodeller. Därefter följer i Kapitel 3 en genomgång hur data har behandlats och vilka verktyg som använts under skrivandet. I Kapitel 4 presenteras datamaterialet och strukturerna på de modeller det har arbetats med. Efter det i Kapitel 5 följer en beskrivning av modellbyggnadsprocessen och prognostiserade värden med vald modell. I Kapitel 6 diskuteras de olika modellerna som arbetats fram. Slutligen följer de referenser som använts samt bilagor.

2. Teoretiskt ramverk

I kapitlet återfinns en beskrivning av de teorier undersökningen kommer att arbeta utifrån när datamaterialet senare ska analyseras. Det presenteras modeller för analys av tidsserier samt antaganden och kriterier för dessa. Kapitlet avslutas med teori om prognoser med hjälp av ARIMA-modeller.

Resonemangen i följande avsnitt förutsätter att läsaren har en del grundkunskaper inom ämnet statistik. För en djupare insikt i tidsserieanalys kan kapitel 1-5 i Chatfield (2004) rekommenderas.

2.1 Tidsserier

En tidsserie är en realisering av en stokastisk process $\{X_t\}$ där sekvensen av observationer utvecklas i tiden enligt sannolikhetslagar och $t = 1, 2, 3, \ldots, T$. Därav finns det till den stokastiska processen en simultan fördelning. Om denna fördelning är känd så är det möjligt att göra uttalanden om sannolika framtida värden på X_t. Ett sätt att göra denna form av prognos är genom så kallade ARIMA(*Autoregressive Integrated Moving Average)* -modeller. (Brockwell och Davis, 2002)

2.2 ARIMA

Box och Jenkins (1970) framförde med sin bok ett par nya slags modeller för att kunna beskriva de processer som ligger till grund för en tidsserie, nämligen gruppen ARMA-modeller: *Autoregressiv-, Glidande medelvärdes- och Autoregressiv-Glidande medelvärdesmodell* som är utformade från av att värdet i en tidsserie kan beskrivas som resultatet av dels tidigare värden och dels en oberoende slumpterm. Om det finns information i historiska observationer, så kallad

tidsförskjuten information som kan hjälpa till att förklara dagens värden kan ARMA-modeller användas. En grundläggande förutsättning vid användandet av dessa modeller är att tidsserien är stationär. (Box och Jenkins, 1970)

2.2.1 Stationaritet

När man använder ARMA-modeller vid prognoser finns det två restriktioner som tillämpas, strikt stationaritet och svag stationaritet. En stokastisk process av slumpmässiga variabler som är indexerade av tiden kan sägas vara strikt stationär om de egenskaper den besitter är opåverkade vid en tidsförändring. Den sammantagna sannolikhetsfördelningen vid tidpunkterna $t_1, t_2, ..., t_T$ ska därmed vara densamma som den sammantagna sannolikhetsfördelningen vid tidpunkterna $t_{1+k}, t_{2+k}, ..., t_{T+k}$ där k representerar tidsförskjutningen, det vill säga ett godtyckligt skift i tiden. Den marginella sannolikhetsfördelningen ändras således inte över tiden, vilket ger att både medelvärdet och variansen måste vara konstant över tiden. Autokovariansen kan ej vara beroende av tidpunkterna utan endast av tidsförskjutningarna mellan två tidspunkter. (Chatfield, 2004)

Strikt stationaritet är ett väldigt restriktivt antagande och därför används svag stationaritet vid praktisk tillämpning. Svag stationaritet är det som fortsättningsvis kommer att refereras till i denna uppsats och definieras som:

1. $E(X_t) = \mu$	konstant för alla t
2. $V(X_t) = \sigma = \gamma_0$	konstant för alla t
3. $COV(X_t, X_{t+k}) = \gamma_k$	autokovariansen är endast beroende av tidsförskjutningen mellan två tidpunkter och inte av tidpunkterna själva

När det råder stationaritet så visar processen upp följande egenskaper:

- Stickprovsmedelvärdet konvergerar mot processens sanna medelvärde
- Variansestimaten konvergerar mot processens sanna varians

Detta är egenskaper som ytterst sällan finns hos icke-stationära processer på grund av att det kan förekomma trend, konjunktursvängningar eller säsongsvariationer bland observationerna.

En icke-stationär process kan differentieras eller säsongdifferentieras för att hamna på en stationär nivå:

$$\Delta X_t = X_t - X_{t-1} \qquad \Delta_s X_t = X_t - X_{t-s}$$

Där Δ är differentieringsoperator och s är säsongsperiodiciteten

I en ARIMA-modell står I för Integrated vilket uttrycker hur många gånger en process behöver differentieras för att hamna på en stationär nivå. (Box och Jenkins, 1970)

2.2.1.1 Dickey-Fuller test om stationaritet

Ett sätt att testa om en tidsserie är stationär är genom ett Dickey-Fuller test om enhetsrot. Man brukar säga att om orginaldatat är icke-stationärt men efter en differentiering är stationärt så förekommer det en enhetsrot (Chatfield, 2004). Dickey-Fuller test under en AR(1)-modell genomförs enligt (Dickey och Fuller, 1979):

AR(1): $X_t = \alpha X_{t-1} + \varepsilon_t$ (se nedan för definition av autoregressiva processer)

Om α = 1 så har man en så kallad *random walk* och icke-stationaritet eftersom medelvärde och varians då beror av t

$$\Delta X = X_t - X_{t-1} = AR(1) - X_{t-1} = \alpha X_{t-1} + \varepsilon_t - X_{t-1}$$

$$\rightarrow \Delta X = (\alpha - 1) X_{t-1} + \varepsilon_t$$

Låt $(\alpha - 1) = \delta$

$$\rightarrow \Delta X = \delta X_{t-1} + \varepsilon_t$$

Testet utgår från följande hypotesuppställning:

$H_0 : \delta = 0$ → Enhetsrot (icke-stationaritet)

$H_A : \delta < 0$ → Stationaritet

Sedan görs en minsta-kvadrat regression genom origo på $\Delta X = \delta X_{t-1} + \varepsilon_t$ för att skatta värdet på δ. Arbetar man med säsongsdata görs regressionen mellan de säsongsdifferentierade observationerna och observationerna med säsongsstor tidsförskjutning: $\Delta_s X = \delta X_{t-s} + \varepsilon_t$. Teststatistikan har följande utseende: (Cheung och Lai, 1995)

$$\text{Teststatistikan: } t = \frac{\hat{\delta}}{s\left(\hat{\delta}\right)}$$

Teststatistikan är inte t-fördelad utan följer en specifik fördelning som brukar refereras som *Dickey-Fuller Table*. Den kritiska nivån för en modell med en parameter ligger på t_{krit} = -1,96 vid en signifikansnivå på 5%.

2.2.2 Purely random process (Vitt Brus)

En process $\{\varepsilon_t\}$ kan definieras som *vitt brus* om de stokastiska variablerna ε_t och ε_{t-k} är okorrelerade, det vill säga: $Cov(\varepsilon_t, \varepsilon_{t-k}) = 0$ där $k > 0$, och om ε_t och ε_{t-k} har samma fördelning. När man har *vitt brus* så råder det ingen autokorrelation mellan slumptermerna. Autokorrelation definieras nedan. Om ε_t och ε_{t-k} är ömsesidigt okorrelerade och normalfördelade med medelvärde 0 och varians σ^2 så brukar det definieras som *Gaussiskt vitt brus.*

2.2.2.1 Ljung-Box test av vitt brus

Ett sätt att kontrollera om det råder *vitt brus* är genom ett Ljung-Box test. Testet är ett slags *overall*-test som prövar modellens totala slumpmässighet och utgår från följande hypoteser (Ljung och Box, 1978):

$H_0 : \rho_1 = \rho_2 = ... = \rho_m = 0$ (*vitt brus)*

H_A : Åtminstone något $\rho_i \neq 0$ i = 1, 2, …, m

Där ρ är autokorrelationen mellan olika tidspunkter och ρ_1 är autokorrelationen mellan ε_t och ε_{t-1}

Se nedan för definition av autokorrelationen

m står för maximal tidsförskjutning

Teststatistikan: LB = $T(T+2)\sum_{k=1}^{m} \frac{r_k^2}{T-k} \sim \chi^2(m)$

Där r_k är den skattade autokorrelationen och T står för totala antalet observationer i tiden

Ljung-Box-statistikan är approximativt chi-två fördelad med m frihetsgrader. Ett problem enligt Chatfield (2004) med Ljung-Box-testet är att det har låg styrka vilket säger att sannolikheten att förkasta nollhypotesen när den är falsk är liten.

2.2.3 Autoregressiv modell (AR)

En AR(p)-process baseras på att en observation i tiden kan förklaras av en linjärkombination av observationer från tidigare tidpunkter. En AR-process av ordning 1 har följande utseende:

$$\text{AR(1): } X_t = \alpha X_{t-1} + \varepsilon_t$$

En AR-process kan utökas till en process av ordning *p* och skrivs då som AR(*p*):

$$\text{AR}(p): X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + ... + \alpha_p X_{t-p} + \varepsilon_t$$

Detta är den generella autoregressiva modellen av ordningen p, där p är antalet parametrar.

ε_t är slumptermen som är en oberoende slumpvariabel med väntevärde 0 och varians som är konstant över tid. Man kan även införa en konstant, μ i ekvationen för att eventuellt uppnå en bättre anpassning av tidsserien. (Baudin, 1996)

2.2.4 Glidande medelvärdes modell (MA)

En MA(q)-process baseras på att en observation i tiden kan förklaras av en linjärkombination av slumptermen vid tidigare tidpunkter. En MA-process av ordningen 1 ser ut som följande:

$$\text{MA(1): } X_t = \beta_1 \varepsilon_{t-1} + \varepsilon_t$$

En MA-process kan även den utökas till en process av ordning q och skrivs då som MA(q):

$$\text{MA}(q):\ X_t = \beta_1\varepsilon_{t-1} + \beta_2\varepsilon_{t-2} + ... + \beta_q\varepsilon_{t-q} + \varepsilon_t$$

Detta är den generella glidande medelvärdesmodellen av ordningen q, där q är antalet parametrar. ε_t är oberoende slumptermer med väntevärde 0 och en varians som är konstant över tid. Även här kan modellen utökas med en konstant, μ för en bättre anpassning. (Baudin, 1996)

2.2.5 Sammansatt ARIMA och SARIMA

Om man kombinerar en MA-process med en AR-process så får man en sammansatt modell som skrivs som $\text{ARMA}(p,q)$ och om man tillför antalet gånger processen behöver differentieras för att bli stationär, d så får man en $\text{ARIMA}(p,d,q)$. Antalet parametrar betecknas av p och q och räknas ut med maximum likelihoodmetoden i program som SPSS och Minitab vilket kommer att användas i följande parameterskattningar. En $\text{ARMA}(p,q)$-process har följande utseende:

$$\text{ARMA}(p,q):\ X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + ... + \alpha_p X_{t-p} + \beta_1\varepsilon_{t-1} + \beta_2\varepsilon_{t-2} + ... + \beta_q\varepsilon_{t-q} + \varepsilon_t$$

(Det går även här inkludera en konstant, μ för en eventuell bättre anpassning)

Box-Jenkins ARIMA-modeller kan även utvidgas till att hantera säsongsvariation. Dessa kallas SARIMA-modeller och den generella tillämpningen är analogt med ARIMA-modeller. Detta ger ARIMA-processen ytterligare en dimension och skrivs som $\text{SARIMA}(p,d,q)(P,D,Q)_s$ där s refererar till säsong. P, D och Q är analoga beteckningar för antalet parametrar och differentieringar (Box och Jenkins, 1970). En säsongs-AR skrivs som SAR(P) och en

säsongs-MA skrivs som SMA(Q). Exempelvis ser en SARIMA $(0,0,0)(1,0,1)$ där tidsserien är månadsindelad ut på följande vis (Chatfield, 2004):

$$\text{SARIMA}\,(0,0,0)(1,0,1)_{s=12}:\ X_t = \alpha_1 X_{t-12} + \beta_1 \varepsilon_{t-12} + \varepsilon_t$$

2.2.6 Autokorrelations- och Partiella autokorrelations-funktionen (ACF & PACF)

Ett bra sätt att identifiera vilken typ av ARIMA-modell man vill använda är att studera autokorrelationsfunktionen (ACF) och den partiella autokorrelationsfunktionen (PACF). Autokorrelationen i en tidsserie beskriver relationen mellan observationer av samma variabel gjorda vid olika tidpunkter, det vill säga korrelationen av variabeln mellan olika tidpunkter. Autokorrelationskoefficienten i tidsförskjutningen *k* definieras som:

$$\rho_k = \frac{COV(X_t, X_{t\pm k})}{V(X_t)} = \frac{\gamma_k}{\gamma_0} \quad (0 \le \rho_k \le 1)$$

Varierar man på *k* så får man ACF. Om man till exempel redan skapat en modell och vill se om det finns kvar någon korrelation mellan slumptermerna så ersätts ovan funktion med ε_t och $\varepsilon_{t\pm k}$. Skattningen av autokorrelationskoefficienten ser ut som följande i tidsförskjutningen *k* med *T* antal observationer: (Brockwell och Davis, 2002)

$$r_k = \frac{\sum_{t=1}^{T-k}(X_t - \bar{X})(X_{t+k} - \bar{X})}{\sum_{t=1}^{T}(X_t - \bar{X})^2} \qquad (\bar{X} \text{ är lika med noll då processen är stationär})$$

PACF däremot ger det marginella bidraget av den senast införda termen i en modell, det vill säga autokorrelationen för den sist inkluderade tidsförskjutningen, givet att de andra tidsförskjutningarna eliminerats ur modellen. Det vill säga autokorrelationen mellan X_t och X_{t-k} när effekten av mellanliggande observationer eliminerats.

För att identifiera p, d och q vid ARIMA-modellering samt även P, D och Q vid SARIMA-modellering kan man genom att studera beteendet hos ACF och PACF, i ett så kallat korrelogram, avgöra vilken nivå på parametrarna och vilken grad av differentiering man bör välja. Beteendet hos ACF och PACF för de olika processerna följer mönstret i Tabell 2.1 nedan (Box och Jenkins, 1970):

Tabell 2.1. *Beteendemönster hos ACF/PACF*

	ACF	PACF
Vitt brus	Ligger nära noll	Ligger nära noll
AR(p)	Exponentiellt avtagande	Spikar till tidsförskjutning p
MA(q)	Spikar till tidsförskjutning q	Exponentiellt avtagande
ARMA(p,q)	Exponentiellt avtagande	Exponentiellt avtagande
SAR$(P)_s$	Exponentiellt avtagande för var s:de tidsförskjutning	Spikar till lag p för var s:de tidsförskjutning
SMA$(Q)_s$	Spikar till lag q för var s:de tidsförskjutning	Exponentiellt avtagande för var s:de tidsförskjutning
SARMA$(P,Q)_s$	Exponentiellt avtagande för var s:de tidsförskjutning	Exponentiellt avtagande för var s:de tidsförskjutning

2.2.7 Informationskriterier

När man arbetar med tidsseriedata som är stationära och vill konstruera en passande ARIMA-modell finns det olika lämplighetsmått för att avgöra hur bra modellens anpassning är till det underliggande datamaterialet. Dessa lämplighetsmått refereras ofta som informationskriterier och är under modellbyggandets gång ett redskap för att jämföra olika modeller, samt ett hjälpmedel för att fastställa bäst passande modell, när det stundom kan vara svårt att tolka ACF och PACF.

2.2.7.1 Akaike's informationskriterium (AIC)

Det vanligaste och det mest använda lämplighetsmåttet som används vid modellering för att avgöra modellens anpassning är *AIC*. Måttet består av två termer, en som baseras på *log-likelihood*-funktionen och en som är ett straff för antalet parametrar i modellen. *AIC* skrivs på följande sätt:

$$AIC : -2(\log - likelihood) + 2g$$

Där g är antalet parametrar: $p + q + P + Q + 1$

(+ 1 kommer från skattningen av residualvariansen)

Värdet på *AIC* i sig är inte användbart utan det är ett mått för att jämföra mellan olika modeller under modellbyggandets gång. Ju lägre värde på *AIC* man har desto bättre förklarar modellen de observationer man har. Som man ser i funktionen så ger ett ökat antal parametrar ett högre värde på *AIC*. Ett problem med *AIC* är att det kan innehålla *bias* om man har få observationer. Därför kan det vara bra att använda ett mått som även tar hänsyn till antalet observationer. (Chatfield, 2004)

2.2.7.2 Normalized Bayesian Information Criterion (Normalized BIC)

Normalized BIC är ett övergripande lämplighetsmått som försöker ta hänsyn till modellens komplexitet. Det baseras på *mean squared error* (MSE) och innehåller ett straff för antalet parametrar i modellen samt för längden på tidsserien. Straffet tar bort fördelen för modeller med fler parametrar, vilket gör att det blir lätt att jämföra olika modeller för samma tidsserie.

Normalized BIC skrivs på följande sätt:

$$\textit{Normalized BIC}:\ \ln(MSE) + g\frac{\ln(T)}{T}$$

Där g är antalet parametrar i modellen, det vill säga $p + q + P + Q$ samt + 1 *om man använt sig av en konstant* μ

T är det totala antalet observationer

Likt *AIC* så är ett lägre värde på *Normalized BIC* en indikation på en bättre anpassad modell till datamaterialet. Värdet i sig säger dock inte något utan det används för att jämföra modeller emellan. Som man ser i formeln så ökar *Normalized BIC* om *g* ökar samt sjunker om *T* ökar. (Dy och Brodley, 2000)

2.2.7.3 Stationary R-squared

Eftersom det vanliga måttet på förklaringsgrad (R^2) endast är lämpligt när det finns minst en förklarande variabel med i ekvationen får man vid utvärdering av tidsseriemodeller använda sig av alternativa mått. Detta på grund av att tidsseriemodeller ofta bara består av tidsförskjutningar och tidigare slumptermer.

Stationary R-squared (Stat. R^2*)* är ett mått där man jämför den nuvarande modellen med en stationär basmodell: ARIMA(0, d, 0)(0, D, 0) (Harvey, 1989)

$$Stat.\,R^2 = 1 - \frac{\sum_t (\varepsilon_t - \hat{\varepsilon}_t)^2}{\sum_t (\Delta\varepsilon_t - \overline{\Delta\varepsilon_t})^2}$$

Stat. R^2 kan anta värden mellan $[-\infty, 1]$.

$Stat.R^2 < 0$ → Den nuvarande modellen är sämre än basmodellen.

$Stat.R^2 = 0$ → Den nuvarande modellen är lika bra som basmodellen.

$Stat.R^2 > 0$ → Den nuvarande modellen är bättre än basmodellen.

2.3 Prognostisering

När man arbetar med prognoser finns det en mängd olika modeller man kan använda. Vad man gör är att skapa en modell, en förenklad bild av verkligheten, och genom den gör man förenklade utsagor om framtiden. Ett vanligt sätt är att använda ARIMA-modeller för prognoser. Prognoser kan göras för en till flera tidsenheter framåt definierat enligt 1, 2, 3, …, *u* steg framåt. I program som SPSS och Minitab kan prognoser skrivas ut efter det man klarlagt beroendestrukturen i ARIMA-modellen. (Chatfield, 2004)

I exempelvis en MA(1) -process ser en prognos för *u* steg framåt ut enligt följande med väntevärde och skattning av prognosfelet:

$$X_{T+u} = \beta_1 \varepsilon_{T+u-1} + \varepsilon_{T+u}$$

$$E(X_{T+u}|X_T) = E(\beta_1 \varepsilon_{T+u-1}|X_T) + E(\varepsilon_{T+u}|X_T) = 0 \text{ eller } = \mu \text{ om man använder en konstant.}$$

$$V(X_{T+u}|X_T) = V(\beta_1 \varepsilon_{T+u-1}|X_T) + V(\varepsilon_{T+u}|X_T) = \sigma^2(1+\beta_1^2)$$

Ett skattat 95%-igt prediktionsintervall ser då ut som följande vid användning av en konstant:

$$\hat{\mu} \pm 1{,}96 \cdot \sqrt{\hat{\sigma}^2(1+\hat{\beta}_1^2)}$$

I en AR(1) -process ser en prognos för *u* steg framåt ut enligt följande med väntevärde och skattning av prognosfelet:

$$X_{T+u} = \alpha X_{T+u-1} + \varepsilon_{T+u}$$

Successiv substitution ger:

$$X_{T+u} = \alpha^u X_T + (\varepsilon_{T+u} + \alpha\varepsilon_{T+u-1} + \alpha^2 \varepsilon_{T+u-2} + \ldots + \alpha^{u-1} \underbrace{\varepsilon_{T+u-(u-1)}}_{T+1})$$

$$E(X_{T+u}|X_T) = E(\alpha^u X_T|X_T) + E(\varepsilon_{T+u} + \alpha\varepsilon_{T+u-1} + \alpha^2\varepsilon_{T+u-2} + \ldots + \alpha^{u-1}\underbrace{\varepsilon_{T+u-(u-1)}}_{T+1})\Big|X_T) = \alpha^u X_T$$

$$V(X_{T+u}|X_T) = V(\alpha^u X_T|X_T) + V(\varepsilon_{T+u} + \alpha\varepsilon_{T+u-1} + \alpha^2\varepsilon_{T+u-2} + \ldots + \alpha^{u-1}\underbrace{\varepsilon_{T+u-(u-1)}}_{T+1})\Big|X_T) =$$

$$= 0 + \sigma^2(1 + \alpha^2 + \alpha^4 + ... + \alpha^{2(u-1)}) = \sigma^2(1 + \alpha^2 + \alpha^4 + ... + \alpha^{2(u-1)})$$

Ett skattat 95%-igt prediktionsintervall ser då ut som följande:

$$\hat{\alpha}^u X_T \pm 1{,}96 \cdot \sqrt{\hat{\sigma}^2(1 + \hat{\alpha}^2 + \hat{\alpha}^4 + ... + \hat{\alpha}^{2(u-1)})}$$

3. Metod

Nedan följer en redogörelse av arbetsgången i undersökningen och de metoder som appliceras på problemet samt dess omfattning och valet av ansats. Därefter följer en beskrivning av datainsamlingsprocessen samt hanteringen av datamaterialet. Kapitlet avslutas med en diskussion kring metodkritik.

3.1 Box-Jenkins modellbyggnadsmetodologi

I Box och Jenkins (1970) presenterades en ny iterativ metod för att konstruera ARMA-modeller. Metoden består av en tre stegs iterativ procedur, identifikation, parameterskattning och utvärdering.

I denna undersökning kommer det att arbetas utifrån en modifierad version av denna metod som även inkluderar förbehandling av datamaterialet samt prognostisering (Makridakis m fl., 1998). Denna fem stegs iterativa metod har följande arbetsgång.

1. Förbehandling: Först kommer det att skattas parametrar genom enkel linjär regression med dummyvariabler för timme, dag eller månad för att se säsongsvariationerna för att sedan gå vidare med ARIMA-modellering och därför behöver datamaterialet uppfylla antagandet om normalitet för att dessa skattningar ska vara korrekta och tolkningsbara. Om det inte råder normalitet bland observationerna kan man transformera datamaterialet för att uppfylla detta krav. Transformeringen kan även hjälpa till med att stabilisera variansen när det råder heteroskedasticitet då variansen varierar över tiden. (Tabachnick och Fidell, 2006)

En förutsättning för ARIMA-modellering är att datamaterialet kan antas uppfylla villkoren för svag stationaritet. Detta kan undersökas genom att rita upp observationerna över tid för att se om det råder någon trend. Det går även att testa genom ett Dickey-Fuller test. Om tidsserien inte är stationär så ska den differentieras. (Box och Jenkins, 1970)

2. Val av modell: Här används grafer för ACF och PACF på det förbehandlade datamaterialet för att kunna identifiera potentiella ARIMA-processer enligt de ovan nämnda identifikationsmönstren. För att testa modellens anpassning bedöms de olika lämplighetsmåtten, även kallade informationskriterier**.**

3. Parameterskattning: Här skattas parametrarna i den ARIMA-modell man valt.

4. Utvärdering av modellen: Här testar man om de krav som satts är uppfyllda som till exempel stationaritetskravet. Man undersöker även om de parametrar man skattat i ARIMA-modellen är signifikanta. För att se om residualerna endast består av *vitt brus* används *Ljung-Box* statistikan.

Steg 1- 4 upprepas till man hittat en acceptabel modell.

5. Prognostisering: Nu är modellen klar och kan användas för att göra prognoser med.

3.2 Val av ansats

Uppsatsen kommer att utföras på ett deduktivt sätt, det vill säga att den utgår från det teoretiska ramverket för att beskriva verkligheten. Olika tidsserier kommer att undersökas genom att använda diverse ARIMA-processer för att modellera dem.

3.3 Urval

De prognosmodeller som kommer itereras fram är en timprismodell, en dagsmedelprismodell och en månadsmedelprismodell.

Eftersom antalet observationer på timpriser låg på över 100 000 och på grund av att man inte behöver så många observationer bakåt i tiden när det gäller korttidsprognoser så kortades antalet observationer ned till enbart statistik från år 2009. Dagsmedel och månadsmedel omfattar alla observationer.

3.4 Datainsamlingen

Allt det datamaterial som uppsatsen behandlar har erhållits från Nord Pool**.** Detta innebär alltså att undersökningen baseras på officiella data, det vill säga sekundärdata. Datamaterialet som erhållits i form av Excel dokument har senare flyttats och använts i statistikprogrammen SPSS och Minitab.

3.5 Metodkritik

3.5.1 Validitet

Validitet säger hur bra en undersökning mäter det den är ämnad att mäta. Då målet är att använda ARIMA-modeller för att mäta skillnader och variationer i elpriser för att senare göra prognoser så är det enda tänkbara studieobjektet elpriser.

3.5.2 Reliabilitet

För att undvika att resultatet påverkas av författaren själv har det använts dataprogram vid alla beräkningar. Det har använts båda statistikprogrammen SPSS och Minitab. Så för att undvika systematiska fel under beräkningarna så har det där det är möjligt beräknats samma sak i båda programmen så man kan se om det finns problem med reliabiliteten hos dataprogrammen. Detta är viktigt för att se så att inte valet av program påverkar resultatet.

3.5.3 Källkritik

När det gäller information från tidigare studier så har det arbetats med flera olika källor i ämnet och jämförts dem sinsemellan för att kunna säkerställa deras trovärdighet. När det gäller all bakgrundsinformation inom Nord Pool så är den hämtad från deras hemsida på internet vilken kanske blir oåtkomlig för framtida läsare då sidan kanske flyttats eller lagts ned.

4. Data

I kapitlet finns en kort genomgång av det datamaterial som tillhandahållits samt strukturen på de olika modellerna som valts att arbeta med.

4.1 Datamaterialet

Det datamaterial som tillhandahållits kommer från NordPool och består av elpriser och är uppdelade i timpriser, dagsmedelpriser, veckomedelpriser samt månadsmedelpriser. Elpriset är det pris som sätts klockan 12.00 varje dag genom att utbud och efterfrågan på elektricitet bland de deltagande på elmarknaden sammanträffar. Datamaterialet sträcker sig från år 1996 till juni 2009.

I denna undersökning kommer det att skapas modeller för timpriser, dagsmedelpriser och för månadsmedelpriser. Detta för att kunna göra både kortsiktiga och långsiktiga prognoser. I de olika modellerna kommer den endogena variabeln motsvara de tre olika priserna och de exogena variablerna motsvarar dummyvariabler för olika typer av säsongsvariationer.

4.1.1 Timprismodellen

I denna modell har det använts tidsseriedata från 2009 bestående av 24 timpriser per dag sträckande från januari till mitten av juni. Med totalt 4008 observationer sträckande över 167 dagar ansågs detta antal vara nog.

Ett problem med en modell med så många observationer är när det senare ska försökas påvisa att ingen korrelation kvarstår. På grund av mängden observationer kan skattade autokorrelationer bli stora endast på grund av slumpen, vilket ger oönskade resultat i *Ljung-Box* testet.

De dummyvariabler, I_i som använts består av påverkan på timpriset beroende på vilken timma på dygnet det är, där $i = 1, 2, 3, \ldots, 24$. Detta görs för att kunna påvisa variationer i priserna inom

dygnet samt för att hjälpa till med att uppfylla antagandet om stationaritet. Senare i uppsatsen kommer vi att se att skattningen av dummyvariablerna gör att den kvarvarande processen kommer kunna betraktas som stationär efter en säsongsdifferentiering vilket den inte hade kunna utan skattningen.

4.1.2 Dagsmedelprismodellen

I denna modell arbetas det med dagsmedelpriser som består av data sträckande från 1996 till juni 2009. Totalt består datamaterialet av 4616 observationer vilket också kan ge problem med test som *Ljung-Box* likt timprismodellen.

De dummyvariabler, I_i som använts i denna modell består av påverkan på dagsmedelpriset beroende på vilken dag det är, där $i = 1, 2, 3, \ldots, 7$. Detta görs för att kunna påvisa variationer i dagsmedelpriset under veckan samt för att hjälpa till med att uppfylla antagandet om stationaritet likt timprismodellen.

4.1.3 Månadsmedelprismodellen

Modellen bygger på 161 observationer på månadsmedelpriser och sträcker sig från 1996 till juni 2009. Då vi har få observationer kan *AIC* vara *biased* men *Normalized BIC* kan användas istället då den tar hänsyn till antalet observationer. (Chatfield, 2004)

De dummyvariabler, I_i som använts i denna modell består av påverkan på månadsmedelpriset beroende på vilket kvartal det är, där $i = 1, 2, 3, 4$. Detta görs för att kunna påvisa variationer i månadsmedelpriset under året samt för att hjälpa till med att uppfylla antagandet om stationaritet likt timprismodellen.

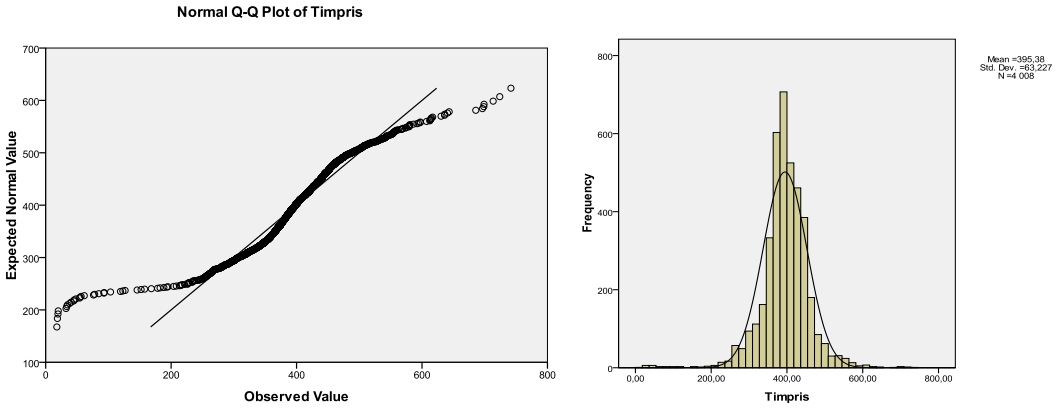
5. Resultat

Nedan följer en redogörelse av modellbyggnadsprocessen för att ta fram adekvata SARIMA-modeller för att beskriva de tre olika tidsserierna som valts att arbeta med. Kapitlet avslutas med prognostisering med hjälp av framtagna modeller. Som hjälpmedel används SPSS och Minitab.

5.1 Timpris

För att avgöra om det är möjligt att arbeta under antagandet om normalitet så ritades observationerna på timpriser i en normalitetsritning samt ett histogram med utritad normalfördelningskurva:

Figur 5.1



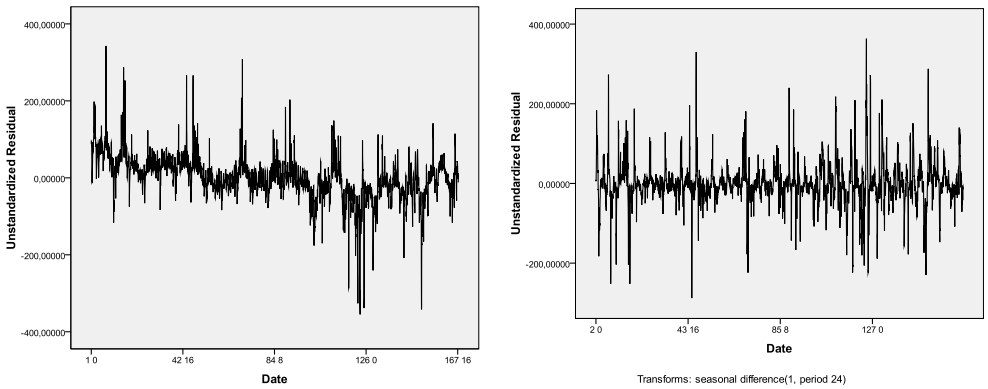
Normalitetsritning (t.v.) och histogram med normalitetskurva (t.h.) över timpriser

Av normalitetsritningen samt histogrammet framgår att en del observationer ligger utanför normalitetskurvan vilket kan bero på outliers. Dessa outliers behålls vid analysen. Datamaterialet anses vara nog anpassat för att arbeta vidare under normalitetsantagandet.

Nästa steg består i att skatta parametrarna för dummyvariablerna med timme 14 som referens. Som synes i Tabell 1.1 i Bilaga 1 så är några av variablerna inte signifikanta vilket gör att man bör vara försiktig med att tolka betydelsen av dessa när man senare ska använda modellen för prognostisering. Residualerna som blir över efter skattningen är de som används för modellering fortsättningsvis.

Efter det undersöks om antagandet om stationaritet är uppfyllt. Detta görs genom att observera ett tidsseriediagram och se om det verkar förekomma någon trend:

Figur 5.2



Tidsseriediagram före säsongsdifferentiering (t.v.) och efter säsongsdifferentiering (t.h.)

Som man ser så verkar det förekomma någon slags nedåtgående trend och eftersom det arbetas med timvärden så säsongsdifferentierades tidsserien med periodiciteten 24.

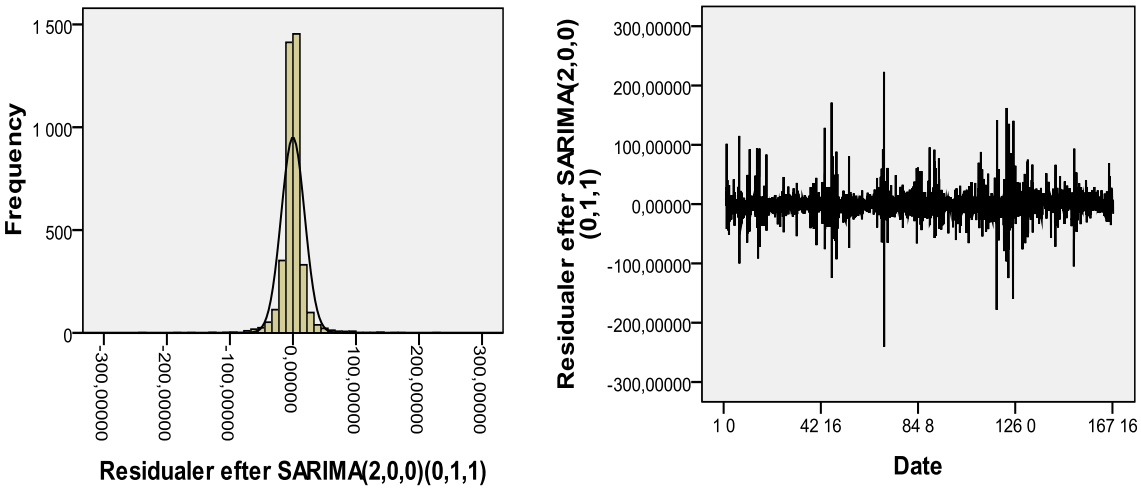
Differentieringen ser ut att ha gjort den resulterande tidsserien stationär vilket sedan testades med hjälp av ett *Dickey-Fuller* test vilket återfinns i Tabell 1.2 i Bilaga 1. Testet gav värdet t_{obs} = -32,08 vilket indikerar på stationaritet då värdet ligger under den kritiska gränsen på -1,96. Därmed kan processen behandlas som stationär och byggandet av en SARIMA-modell kan sätta igång.

För att skapa en SARIMA-modell används ACF och PACF som ritades upp i korrelogram för att identifiera beroendestrukturen. Korrelogrammen plus informationskriterier för samtliga timprismodeller återfinns i Bilaga 1. Korrelogrammen över residualerna efter att dummyvariablerna skattats visar också på att det råder säsongsvariation vid tidsförskjutning 24, 48, 72 vilket även det indikerar på att processen behöver säsongsdifferentieras. Efter säsongsdifferentieringen uppvisar ACF ett exponentiellt avtagande mönster och PACF uppvisar två spikar. Därmed ansattes en $\text{SARIMA}\,(2,0,0)(0,1,0)_{s=24}$. Efter ansättandet av denna modell uppvisar ACF en spik vid 24:e tidsförskjutningen och PACF uppvisar ett avtagande mönster genom var 24:e tidsförskjutning, vid 24, 48, 72 och så vidare. Detta tyder på ett långsiktigt beroende vid den 24:e tidsförskjutningen, vilket kan anses vara ganska förståeligt då elpriset vid en viss timma samvarierar med elpriset vid samma tidpunkt nästa dag. Därmed ansattes en $\text{SARIMA}\,(2,0,0)(0,1,1)_{s=24}$ och när ACF och PACF ritats upp i korrelogram så kunde inget speciellt mönster eller beroende uttydas.

Stat. R^2 har ökat från $-2{,}2 \cdot 10^{-16}$ i modell $\text{SARIMA}\,(0,0,0)(0,1,0)_{s=24}$ till 0,860 i modell $\text{SARIMA}\,(2,0,0)(0,1,1)_{s=24}$. *AIC* har mellan samma modeller minskat från 42452 till 34631 och *Normalized BIC* har minskat från 7,82 till 5,86. Parameterskattningarna för SARIMA-modellen är alla signifikanta. Alla modellmått återfinns i Tabell 1.4. i Bilaga 1.

För att se om det kvarstår *vitt brus* i modellen så studeras *Ljung-Box* testet. Tyvärr förkastas nollhypotesen vilket indikerar på att det inte råder *vitt brus*. Däremot har *Ljung–Box* statistikan minskat från 10603 i modell $\text{SARIMA}\,(0,0,0)(0,1,0)_{s=24}$ till 65,7 i modell $\text{SARIMA}\,(2,0,0)(0,1,1)_{s=24}$ vilket är en rejäl förbättring. I realiteten kan det nämligen vara väldigt svårt att modellera alla beroenden och på grund av det stora antalet observationer kan det finnas ”falska” beroenden som uppkommer endast på grund av slumpen. För att undersöka saken närmare så ritades residualerna upp i ett histogram samt mot tiden för att se om de var normalfördelade och för att se om det kunde uttydas något beroende, som enligt *Ljung-Box* fortfarande föreligger:

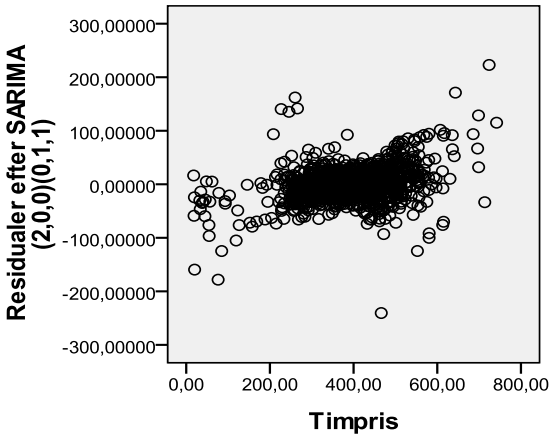
Figur 5.3



Kvarvarande residualer efter modell SARIMA $(2,0,0)(0,1,1)_{s=24}$ *ritat i histogram (t.v.) och mot tiden (t.h.)*

Som man kan se så verkar residualerna vara normalfördelade och man kan inte uttyda något beroende. För att undersöka detta vidare ritades residualerna mot timpriset för att se om det råder heteroskedasticitet:

Figur 5.4



Kvarvarande residualer ritade mot timpris

Man skulle kunna tänka sig att utseendet på föregående figur ger ett intryck av beroende men här anses de vara nog spridda för att anta att det råder homoskedasticitet.

Så med den sammantagna informationen med signifikanta parameterskattningar, otydligt beroendemönster i ACF och PACF, höga värdet på *Stat.* R^2 och de låga värdena på *AIC* och *Normalized BIC* anses modellen vara klar.

Den sammansatta modellen med dummyvariablerna har då följande utseende:

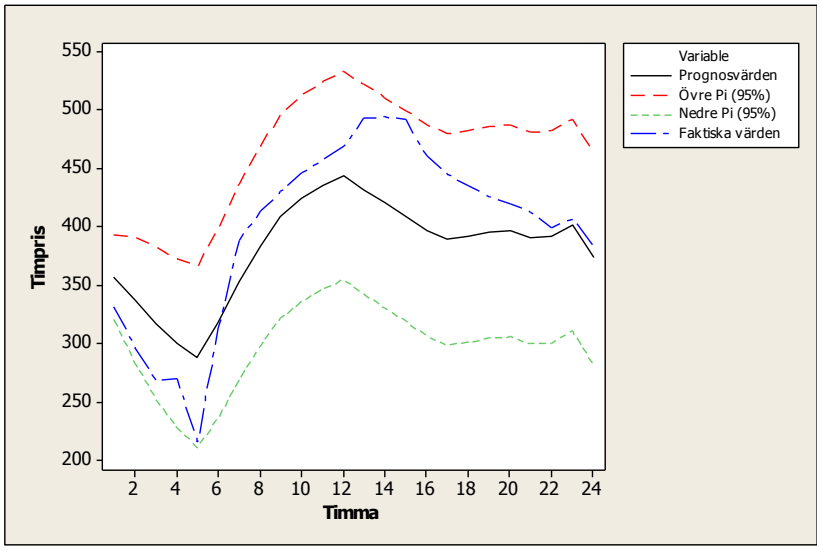
Timpris (X_t) =

$$\alpha + \gamma_i I_i + SARIMA(2,0,0)(0,1,1)_{s=24} =$$
$$410{,}6 - 34{,}8I_1 - 52{,}3I_2 - 62{,}5I_3 - 68{,}9I_4 - 66{,}7I_5 - 50{,}1I_6 - 29{,}68I_7 - 1{,}5I_8 + 22{,}1I_9 + 22{,}4I_{10}$$
$$+ 22{,}2I_{11} + 16{,}5I_{12} + 8{,}4I_{13} - 8{,}7I_{15} - 13{,}4I_{16} - 11{,}6I_{17} - 1{,}1I_{18} + 4{,}5I_{19} - 2{,}0I_{20} - 7{,}2I_{21} - 9{,}6I_{22}$$
$$- 12{,}0I_{23} - 30{,}0I_{24} + X_{t-24} - 0{,}478 + 1{,}099(X_{t-1} - X_{t-25}) - 0{,}204(X_{t-2} - X_{t-26}) + 0{,}88\varepsilon_{t-24} + \varepsilon_t$$

Där I_1 *är indikator för timme 1,* I_2 *indikator för timme 2 o.s.v.*

En prognostisering med ovanstående modell på 24 timmar den 17e juni gav följande utseende:

Figur 5.5

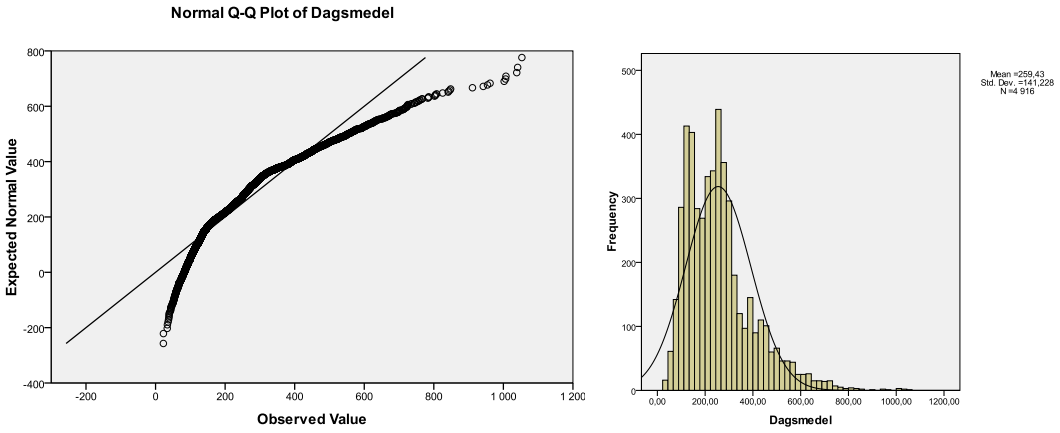


Prognostiserade värden och faktiska värden för timpriser 24 timmar framåt med 95%igt prediktionsintervall. Prognosvärdena återfinns i Tabell 1.5. i Bilaga 1.

5.2 Dagsmedelpris

Byggandet av denna modell blir snarlik den vid timpriser och till en början undersöktes om det gick att arbeta under antagandet om normalitet. Även här ritades observationerna i en normalitetsritning samt i ett histogram med normalfördelningskurva:

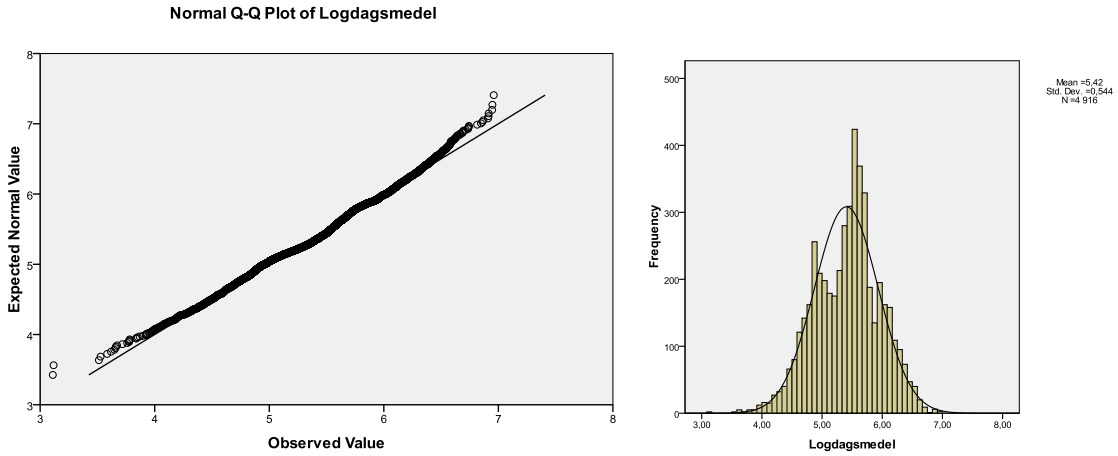
Figur 5.6



Normalitetsritning (t.v.) och histogram med normalitetskurva (t.h.) över dagsmedelpriser

Som man ser i histogrammet så är observationerna skeva åt höger vilket även syns i normalitetsritningen. Detta indikerar att en logaritmisk transformering av variabeln skulle resultera i en bättre anpassning till normalfördelning. En sådan transformation skulle även kunna hjälpa till med att stabilisera en eventuell heteroskedastisk varians. Därmed genomfördes en logaritmisk transformation:

Figur 5.7



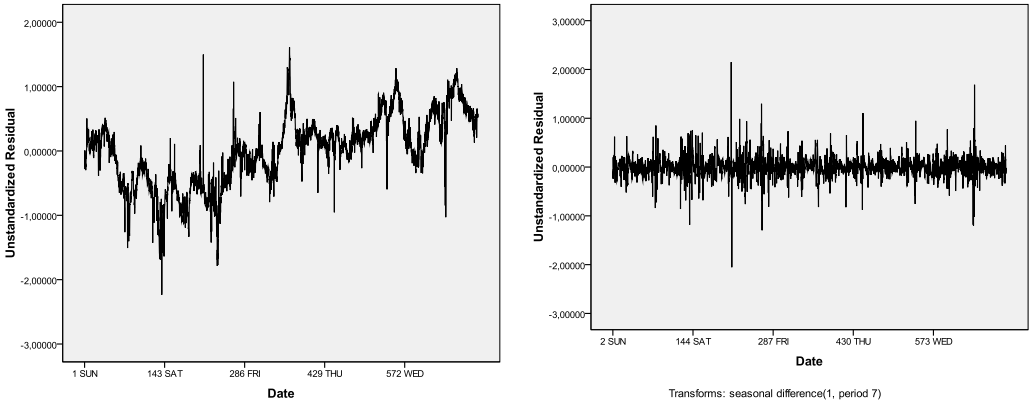
Normalitetsritning (t.v.) och histogram med normalitetskurva (t.h.) över logaritmerade dagsmedelpriser

Som man kan se i figurerna så verkar transformationen möjliggjort att fortsatt arbete med modellen kan ske under ett normalitetsantagande och därför används det transformerade datamaterialet fortsättningsvis.

Därefter skattades parametervärdena för dummyvariablerna med måndag som referensvariabel. Som man kan se i Tabell 2.1 i Bilaga 2 så är det bara lördag, söndag och måndag som visar signifikanta skillnader i prisnivåerna. Man kan alltså bevisa att det finns signifikanta skillnader i dagsmedelpriserna mellan lördag och måndagen. Resterande variabler får vara kvar men bör tolkas med viss försiktighet när det senare ska göras prognoser. Anledningen till att dessa variabler behålls i modellen är att stationaritetstesterna som senare görs uppvisar stationaritet endast då alla variabler är med i modellen. Detta steg kan alltså ses som ett sätt att förprocessera datamaterialet för att uppfylla stationaritetskravet. Residualerna som blir över efter skattningen är de som används för modellering fortsättningsvis.

Nu för att undersöka om observationerna följer ett stationärt utseende används tidsseriediagram för att se om det finns någon trend:

Figur 5.8



Tidsseriediagram före säsongsdifferentiering (t.v.) och efter säsongsdifferentiering (t.h.) för resterande residualer

Som man kan se verkar det råda en uppåtgående trend och eftersom att det är dagsmedelvärden som behandlas så säsongsdifferentierades tidsserien med periodiciteten 7.

Säsongsdifferentieringen verkar ha medfört att observationerna kan betraktas som stationära. För att testa detta används *Dickey-Fuller* test vars resultat återfinns i Tabell 2.2 i Bilaga 2. Testet gav ett resultat på t_{obs} = -11,2 vilket är under den kritiska gränsen på -1,96. Därmed kan processen behandlas som stationär vilket möjliggör sammansättningen av SARIMA-modellen.

För att skapa en SARIMA-modell används även här ACF och PACF som ritades upp i korrelogram för att identifiera beroendestrukturen. Samtliga korrelogram och informationskriterier för dagsmedelprismodellerna återfinns i Bilaga 2. Korrelogrammen över residualerna efter att dummyvariablerna skattats visar på en liten säsongsvariation vid tidsförskjutning 7, 14, 21, 28 vilket indikerar på att processen behöver säsongsdifferentieras. Efter säsongsdifferentieringen uppvisar ACF ett exponentiellt avtagande mönster och PACF uppvisar en spik. Därmed ansattes en $\text{SARIMA}(1,0,0)(0,1,0)_{s=7}$. Efter att denna modell ansatts uppvisar ACF en spik vid 7:e tidsförskjutningen och PACF uppvisar ett avtagande mönster genom var 7:e förskjutning, vid 7, 14, 21, 28 och så vidare. Detta tyder på ett långsiktigt

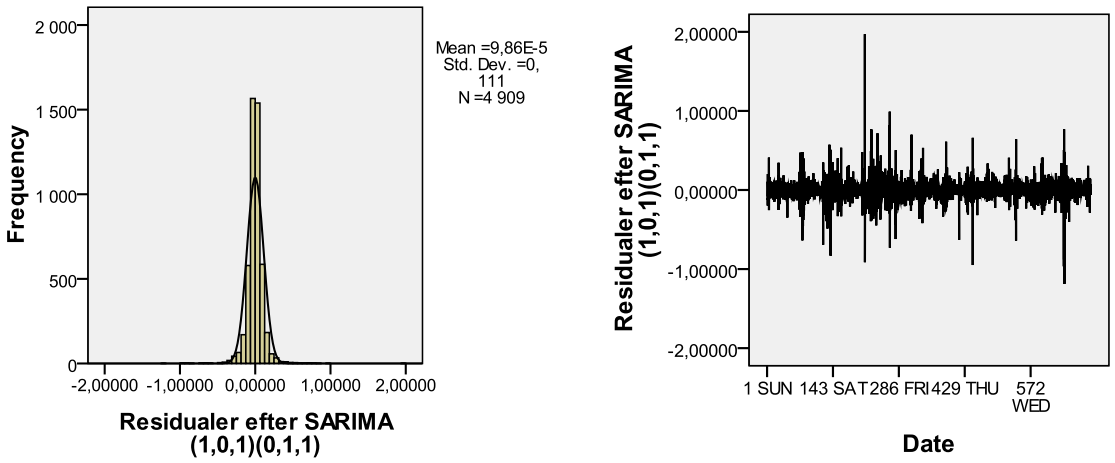
beroende vid den 7:e tidsförskjutningen, då man kan anta att medelpriset är beroende veckovis vilket är ganska förståeligt.

Därmed ansattes en SARIMA $(1,0,0)(0,1,1)_{s=7}$. I denna modell uppvisar ACF en liten spik och PACF ett avtagande mönster så därför provades en SARIMA $(1,0,1)(0,1,1)_{s=7}$ för att se om modellen blir bättre. Nu går inget speciellt mönster att tyda.

Stat. R^2 har ökat från -2,69 $\cdot 10^{-15}$ i modell SARIMA $(0,0,0)(0,1,0)_{s=7}$ till 0,415 i modell SARIMA $(1,0,1)(0,1,1)_{s=7}$. *AIC* har mellan samma modeller minskat från -2754 till -7612 och *Normalized BIC* har minskat från -3,397 till -4,383 och parameterskattningarna för SARIMA-modellen är alla signifikanta. Alla modellmått återfinns i Tabell 2.4. i Bilaga 2.

För att se om det kvarstår *vitt brus* i modellen så studerades *Ljung-Box* testet. Tyvärr kan nollhypotesen förkastas även här vilket indikerar på att det inte råder *vitt brus*. Däremot har *Ljung–Box* statistikan minskat från 4610 i modell SARIMA $(0,0,0)(0,1,0)_{s=24}$ till 120 i modell SARIMA $(1,0,1)(0,1,1)_{s=7}$ vilket är en rejäl förbättring. Även här kan detta bero på svårigheten att modellera alla beroenden som kan uppkomma på grund av det stora antalet observationer och slumpen. Men för att undersöka saken närmare så ritades residualerna upp i ett histogram samt mot tiden för att se om de var normalfördelade och för att se om det kunde uttydas något beroende, som enligt *Ljung-Box* fortfarande föreligger:

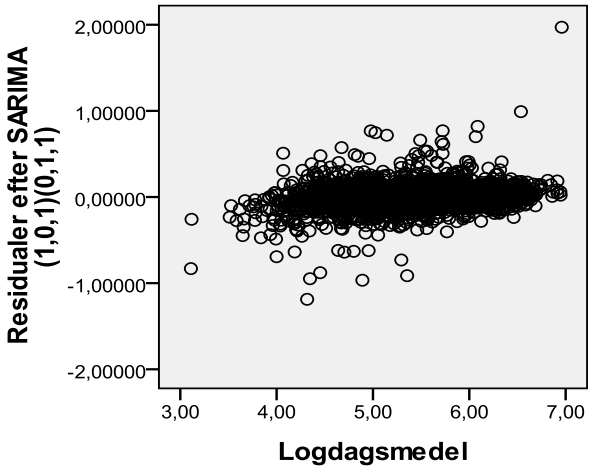
Figur 5.9



Kvarvarande residualer efter modell SARIMA $(1,0,1)(0,1,1)_{s=7}$ *ritat i histogram (t.v.) och mot tiden (t.h.)*

Som man kan se så verkar residualerna vara normalfördelade och man kan inte uttyda något beroende. För att undersöka detta vidare ritades residualerna mot timpriset för att se om det råder heteroskedasticitet:

Figur 5.10



Kvarvarande residualer ritade mot logaritmerat dagsmedelpris

Man skulle även här kunna tänka sig att utseendet på föregående figur ger ett intryck av beroende men anses vara nog spridda för att anta att det råder homoskedasticitet.

Så med den sammantagna informationen med signifikanta parameterskattningar, otydligt beroendemönster i ACF och PACF, höga värdet på $Stat.\,R^2$ och de låga värdena på *AIC* och *Normalized BIC* anses modellen vara klar.

Den sammansatta modellen med dummyvariablerna har då följande utseende:

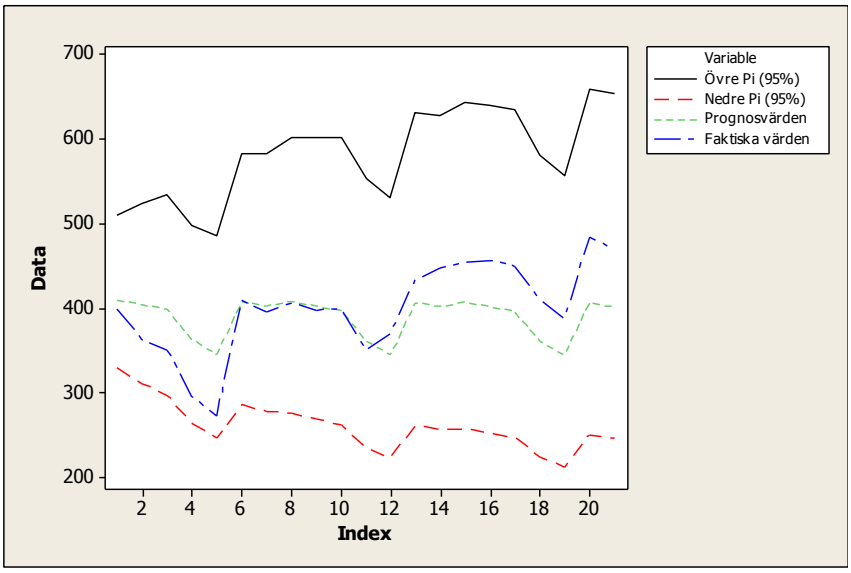
Dagsmedelpris (X_t) = $e^{\alpha+\gamma_i I_i+SARIMA(1,0,1)(0,1,1)_{s=7}} =$

$$e^{\begin{matrix}5,461+0,006I_2+0,002I_3-0,01I_4-0,033I_5-0,115I_6-0,161I_7\\ +X_{t-7}+0,001+0,932(X_{t-1}-X_{t-8})+0,275\varepsilon_{t-1}+0,874\varepsilon_{t-7}+0,24035\varepsilon_{t-8}+\varepsilon_t\end{matrix}}$$

Där I_2 *är indikator för tisdag och* I_3 *är indikator för onsdag o.s.v.*

En prognostisering med ovanstående modell på tre veckor framåt räknat från den onsdagen den 17e juni 2009 gav följande utseende:

Figur 5.11

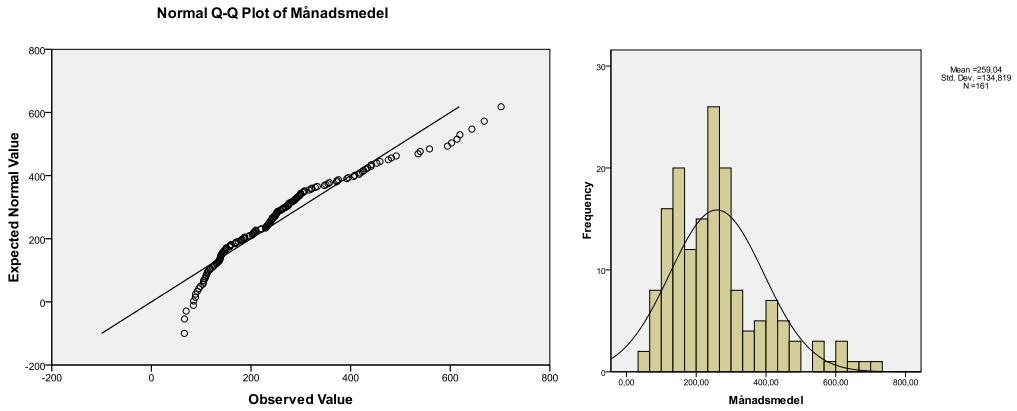


Prognostiserade värden och faktiska värden för dagsmedelpriser 21 dagar framåt med 95%igt prediktionsintervall. Prognosvärdena återfinns i Tabell 2.5. i Bilaga 2.

5.3 Månadsmedelpris

Modellbyggandet för denna tidsserie följer i samma spår som föregående modeller. Till en början kontrollerades om det råder normalitet hos observationerna. Detta görs genom att rita upp observationerna på månadsmedelpriser i en normalitetsritning och ett histogram med normalfördelningskurva:

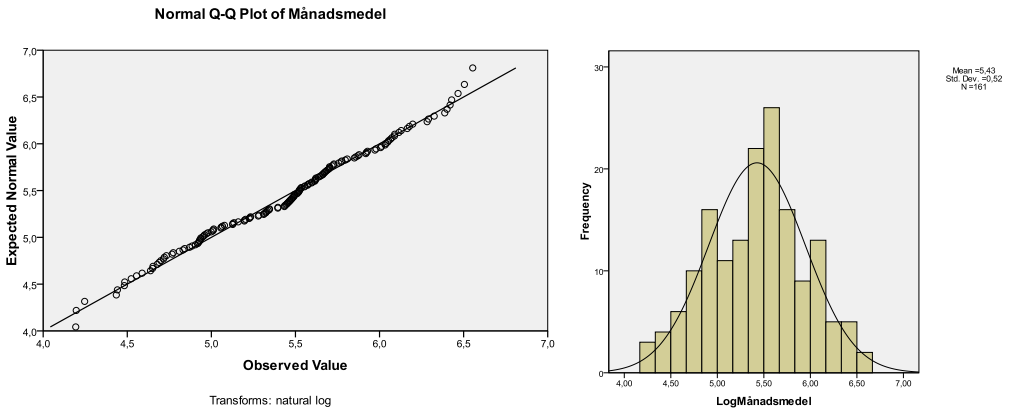
Figur 5.12



Normalitetsritning (t.v.) och histogram med normalitetskurva (t.h.) över månadsmedelpriser

Likt observationerna vid dagsmedelpriserna är dessa också skeva åt höger. Normalitetsritningen indikerar att en logaritmisk tranformering kan hjälpa till att skapa en bättre anpassning till ett normalitetsantagande, samt hjälpa till med att stabilisera eventuell heteroskedastisk varians. Det genomfördes således en logaritmisk transformering:

Figur 5.13



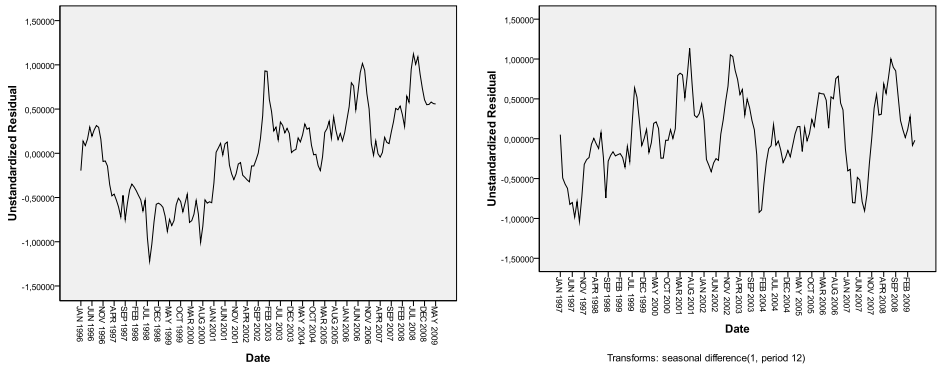
Normalitetsritning (t.v.) och histogram med normalitetskurva (t.h.) över logaritmerade månadsmedelpriser

Man kan se att transformationen har bearbetat observationerna så att arbetet kan ske under antagandet om normalitet.

Därefter skattades parametervärdena för dummyvariablerna med kvartal ett som referensvariabel. Som man kan se i Tabell 3.1 i Bilaga 3 så är ingen av parameterskattningarna signifikanta men som tidigare nämnts behålls så behålls de i modellen för att uppfylla kravet på stationaritet. Betydelsen av regressionsvariablerna bör dock tolkas med viss försiktighet vid senare prognostisering. Koefficienterna är som synes väldigt små och därmed påverkar de inte prognoserna så mycket. De resterande residualer som blir kvar efter skattningen är de som används fortsättningsvis.

För att se om det föreligger någon trend bland observationerna och därmed inte följer ett stationärt utseende skapades tidsseriediagram:

Figur 5.14



Tidsseriediagram före säsongsdifferentiering (t.v.) och efter säsongsdifferentiering (t.h.) för resterande residualer

Av diagrammen kan man uttyda en uppåtgående trend och eftersom det arbetas med månadsmedelvärden så säsongsdifferentierades tidsserien med periodiciteten 12.

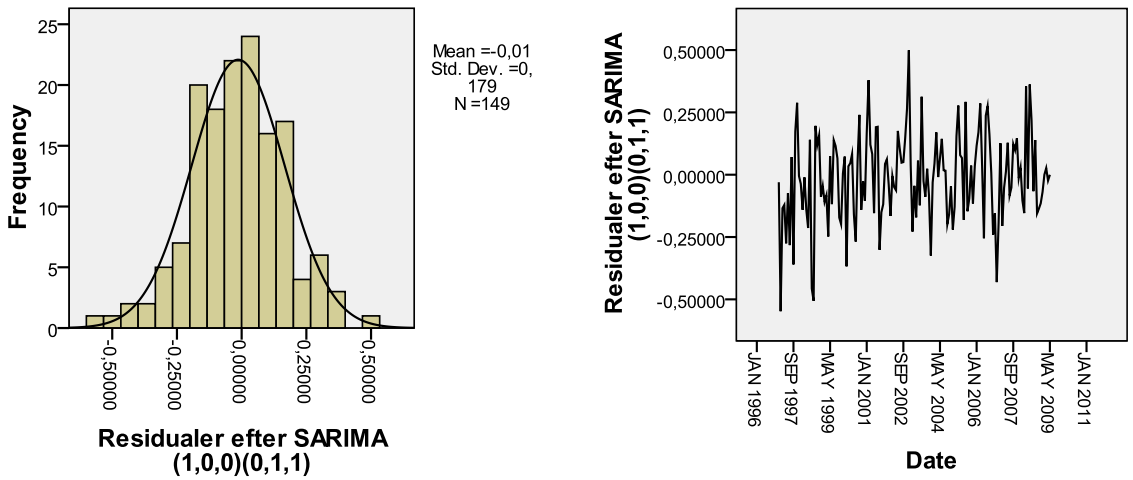
Säsongsdifferentieringen verkar ha medfört att observationerna kan betraktas som stationära vilket kontrollerades med *Dickey-Fuller* test vars resultat finns i Tabell 3.2 i Bilaga 3. Teststatistikan gav värdet t_{obs} = -5,23 vilket är under den kritiska gränsen -1,96 och därmed kan processen behandlas som stationär vilket möjliggör sammansättningen av SARIMA-modellen.

För att identifiera beroendestrukturen ritades ACF och PACF upp i korrelogram. Korrelogrammen och informationskriterierna för månadsmedelprismodellerna återfinns i Bilaga 3. Korrelogrammen över residualerna efter att dummyvariablerna skattats visar på att säsongsdifferentiering bör användas. Efter säsongsdifferentieringen uppvisar ACF ett avtagande mönster och PACF uppvisar en spik. Därför ansattes en $\text{SARIMA}(1,0,0)(0,1,0)_{s=12}$. ACF uppvisade nu en spik vid 12:e tidsförskjutningen och PACF uppvisar ett långsiktigt beroende med ett avtagande mönster genom tidsförskjutning 12, 24, 36 och så vidare. Därmed ansattes en $\text{SARIMA}(1,0,0)(0,1,1)_{s=12}$ och ACF och PACF ritades upp i korrelogram som inte visade något beroendemönster. Därför används informationskriterierna och de olika testen för att undersöka om modellen är signifikant som helhet. Informationskriterierna återfinns i Tabell 3.4. i Bilaga 3.

Ljung-Box testet visar att nollhypotesen om *vitt brus* ej går att förkasta. *Stat.* R^2 har ökat från $-3{,}405 \cdot 10^{-16}$ i modell SARIMA $(0,0,0)(0,1,0)_{s=12}$ till 0,858 i modell SARIMA $(1,0,0)(0,1,1)_{s=12}$. *AIC* har mellan samma modeller minskat från 204 till -85 och *Normalized BIC* har minskat från -1,4 till -3,3. Parameterskattningarna för SARIMA-modellen är alla signifikanta.

På grund av tvivelaktigheten med *Ljung-Box* testet så ritades residualerna upp i ett histogram samt mot tiden för att se om de var normalfördelade och för att se om det kundes uttyda något beroendemönster.

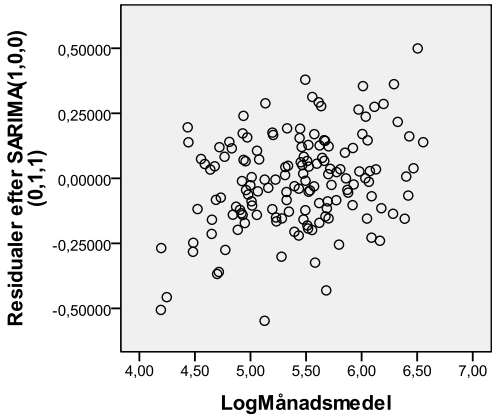
Figur 5.15



Kvarvarande residualer efter modell SARIMA $(1,0,0)(0,1,1)_{s=12}$ *ritat i histogram (t.v.) och mot tiden (t.h.)*

Som man kan se så verkar residualerna vara normalfördelade och man kan inte uttyda något beroendemönster. För att undersöka detta vidare ritades residualerna mot timpriset för att se om det råder heteroskedasticitet:

Figur 5.16



Kvarvarande residualer ritade mot logaritmerat dagsmedelpris

Här ser man ett tydligt tecken på homoskedasticitet.

Så med den sammantagna informationen med signifikanta parameterskattningar, otydligt beroendemönster i ACF och PACF, höga värdet på *Stat.* R^2 och de låga värdena på *AIC* och *Normalized BIC* anses modellen vara klar.

Den sammansatta modellen med dummyvariablerna har då följande utseende:

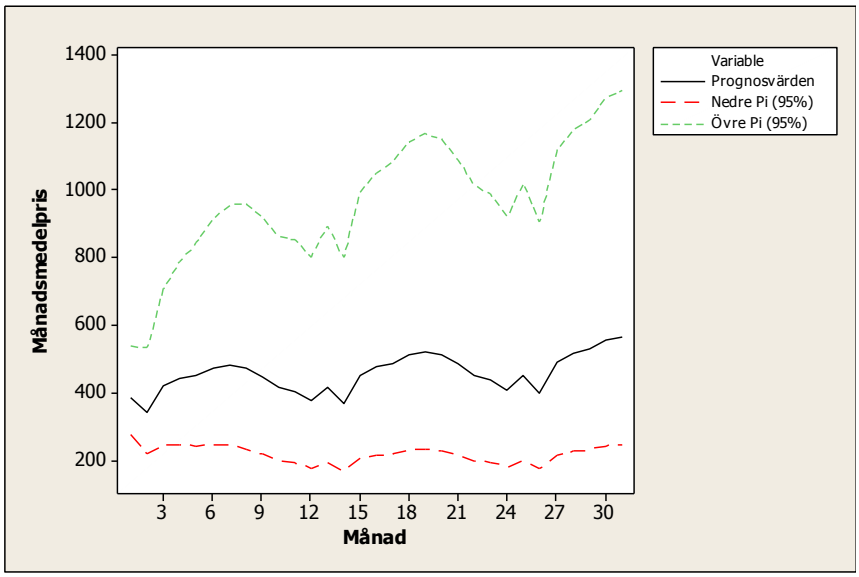
Månadsmedelpris (X_t **) =** $e^{\alpha+\gamma_i I_i+SARIMA(1,0,0)(0,1,1)_{s=12}} =$

$$e^{5,474-0,138I_2-0,114I_3+0,064I_4+X_{t-12}+0,082+091(X_{t-1}-X_{t-13})+0,925\varepsilon_{t-12}+\varepsilon_t}$$

Där I_2 *är indikator för andra kvartalet och* I_3 *är indikator för tredje kvartalet och* I_4 *är indikator för fjärde kvartalet.*

En prognostisering med ovan modell på 31 månader framåt räknat från juni 2009 gav följande utseende:

Figur 5.17



Prognostiserade värden för månadsmedelpriser 31 månader framåt med 95%igt prediktionsintervall. Prognosvärdena återfinns i Tabell 3.5. i Bilaga 3.

6. Diskussion och slutsatser

Uppsatsen avslutas här med diskussion kring resultatet av framtagna modeller och dess prognoser samt exempel på förbättringar och idéer för fortsatt undersökning.

I denna uppsats har tre olika tidsseriemodeller tagits fram med elpriser som underliggande datamaterial för att kunna påvisa tänkbara prisfluktuationer samt för att förutse kommande elpriser. Datamaterialet har erhållits från den nordiska elbörsen Nord Pool vilka agerar som en marknadsplats där elektricitet köps och säljs likt en börsmarknad. De modeller som arbetats med behandlar timpriser, dagsmedelpriser samt månadsmedelpriser och innefattar dummyvariabler som beskriver förändringar i priserna beroende på timma, dag eller kvartal för respektive modell.

När det gäller variationer i priserna ser man i modellen för timpriser att priserna är signifikant lägre under natten mellan ett och sex för att sedan öka vid sjutiden på morgonen fram till tolvtiden för att sedan sjunka mot aftonen. Med andra ord finns det belägg för en ökning av elpriserna under arbetsdagen då det kan tänkas att efterfrågan på elektricitet ökar.

Om man undersöker variationer under veckan så visar modellen för dagsmedelpriserna att medelpriserna tenderar på att vara högre under arbetsdagarna för att sedan sjunka signifikant på helger. Även här finns det belägg för ett lägre elpris under helger då det är tänkbart att efterfrågan på elektricitet sjunker.

Månadsmedelprismodellen visar att medelpriserna tenderar att vara lägre under vår och sommar, det vill säga under kvartal två och tre, för att sedan vara högre under senhösten och vintern. Detta ger då belägg för ett lägre elpris under sommar och vår då det är tänkbart att efterfrågan på elektricitet sjunker på grund av den ökande värmen. Variablerna visar här ingen signifikant skillnad men modellen ger ändå stöd åt den tänkbara skillnaden i elpriserna mellan kvartalen. Dock bör man med viss försiktighet tolka dessa skillnader. Man kan tänka sig att prisskillnaderna har förändrats med åren och med tanke att modellen innefattade observationer från alla år som

Nord Pool varit aktivt så visar parameterskattningarna icke signifikanta skillnader då de ändrats med åren.

De icke signifikanta variablerna har trots sina brister lämnats kvar i modellerna då variabelskattningen har sett till att observationerna kan betraktas som stationära, vilket är väsentligare när man arbetar med prognosmodellering i form av ARIMA-modeller.

Prognoserna för timpriser följer nära de sanna värdena vilket tyder på att prognosmodellen verkar kunna förklara framtida värden. Priserna förväntas minska under natten för att sedan öka under dagen och sedan sjunka mot kvällen vilket de faktiskt gjort. Topparna och dalarna följer inte riktigt det sanna priset men har ett svängningsmönster som ligger nära. På grund av att det använts observationer från hela 2009 så kanske inte modellen fångar upp den senaste tidens extrema värden lika bra som den kanske gjort med färre observationer. Samtliga verkliga värden ligger dock innanför prediktionsintervallet vilket även det tyder på en bra prognosmodell. Man ser även här i prognoserna att prisvariationerna följer de tänkbara fluktuationerna med lägre priser under natten för att senare öka under dagen och sedan sjunka mot kvällen.

Prognoserna för dagsmedelpriser ligger nära det sanna priset likt timprisprognoserna. Men även här når inte prognoserna ut i de extrema värdena. Även här kan det höga antalet observationer försämra modellens förmåga att fånga upp extrema värden. De tänkbara prisvariationerna går även här att skönja med högre priser under veckan som sedan sjunker under helgerna.

Prognoserna för månadsmedelpriserna bevisar även den de tänkbara prisfluktuationerna med lägre priser under sommaren som sedan ökar under vintern. Dessa värden går dock inte att jämföra med några verkliga värden men fungerar bra som en långtidsprognos framåt.

När det gäller modellerna finns det en del svagheter. Till exempel så verkar timprismodellen lida av toppighet vilket skulle kunna bero på outliers. Vad man skulle kunna göra är att beskriva observationerna med någon annan sannolikhetsfördelning eller bara gå igenom datamaterialet och bearbeta bort alla extrema värden. Även *Ljung-Box* testen gav svaga resultat under timpris- och dagsmedelprismodellerna men på grund av att testet anses vara rätt svagt samt det stora antalet observationer har det inte lagts allt för stor tyngd kring det. När man använder höga antal observationer kan det finnas många slumpmässiga beroenden som i realiteten helt enkelt inte går att modellera. Man skulle även kunna påstå i dessa två modeller att det finns antydan till

heteroskedasticitet. För att kunna uppnå homoskedasticitet skulle man kunna införa en så kallad ARCH-modell. Vad denna gör är att ansätta en modell för variansen som gör att den blir homoskedastisk.

Det har alltså kunnats förutspå framtida elpriser som har legat väldigt nära de sanna priserna när det gäller timpriser samt dagsmedelpriser. När det gäller månadsmedelpriser återstår det att se hur bra modellens prognoser har legat de priser som uppstått. Det har även bevisats att de olika prisfluktuationerna som antas finnas existerar.

Klart står alltså att modellerna har fungerat bra som ett hjälpmedel för att förutspå framtida elpriser och för att förklara elprisernas variationer under dagen, veckan samt året. Modellerna skulle dock kunna vidareutvecklas för framtida undersökningar.

Referenser

Vetenskapliga artiklar:

Byström, H.N.E. (2000). *The Hedging Performance of Electricity Futures on the Nordic Power Exchange Nord Pool.* Department of Economics, School of Economics and Management, University of Lund. Working paper series 2000:15

Cheung, Y.W., Lai, K.S. (1995) *Lag order and Critical values of the Augmented Dickey-Fuller Test.* Journal of Business & Ecinomic Statistics, July 1995, Vol 13, No. 3.

Dy, J. G., Brodley, C. E. (2000). *Feature subset selection and order identification for unsupervised learning.* Proc. 17th International Conf. on Machine Learning. Morgan Kaufmann, San Francisco, CA.

Dickey, D.A., Fuller, W.A. (1979). *Distribution of the Estimators for Autoregressive Time Series with a Unit Root.* Journal of the American Statistical Association, 74, p. 427-431

Heden, H. (2007). *Åtskillnaden mellan handel med el och produktion av el: En rapport från Energimarknadsinspektionen.* Statens energimyndighet EMIR 2007:04

Hjalmarsson, E. (2000). *Nord Pool: A Power Market without Market Power*. Department of Economics, Göteborg University. Working papers in economics No 28

Landsorganisationen I Sverige (2007). *Energi för jobb, välfärd och miljö.* Energipolitiskt program för LO

Ljung, G.M., Box, G.E.P. (1978). *On a measure of lack of fit in time series models.* Biometrica **65**, 297-303

Böcker:

Box, G.E.P., Jenkins, G.M. (1970) *Time series analysis: Forecasting and control.* San Francisco: Holden-Day.

Brockwell, P.J., Davis, R.A. (2002) *Introduction to Time Series and Forecasting.* New York: Springer

Chatfield, C. (2004) *The Analysis of Time Series: An Introduction.* Boca Raton: Chapman & Hall/CRC

Harvey, A. C. (1989) *Forecasting, structural time series models and the Kalman filter*, Cambridge University Press.

Makridakis, S., Wheelwright, S.C., Hyndman, R.J. (1998) *Forecasting: methods an applications.* New York: John Wiley & Sons

Tabachnick, B.G., Fidell, L.S. (2006) *Using Multivariate Statistics.* Boston, MA: Allyn & Bacon

Internet:

http://www.nordpool.com, hämtat 2009-06-16

Kompendium:

Baudin, A. (1996). *Prognoser*. Sverige: Statistiska institutionen, Umeå Universitet

, hämtat 2009-06-16

Bilagor

Bilaga 1. Modellbyggnadsprocessen för timpris

Tabell 1.1. *Modellsummering* $X_t = \alpha + \gamma_i I_i$

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,437[a]	,191	,186	57,04326

a. Predictors: (Constant), Timma24, Timma23, Timma22, Timma21, Timma20, Timma19, Timma18, Timma17, Timma16, Timma12, Timma11, Timma10, Timma9, Timma8, Timma7, Timma6, Timma5, Timma4, Timma3, Timma2, Timma1, Timma15, Timma13

b. Dependent Variable: Timpris

Coefficients[a]

Model		Unstandardized Coefficients		t	Sig.
		B	Std. Error		
1	(Constant)	410,632	4,414	93,026	,000
	Timma1	-34,833	6,243	-5,580	,000
	Timma2	-52,252	6,243	-8,370	,000
	Timma3	-62,522	6,243	-10,015	,000
	Timma4	-68,882	6,243	-11,034	,000

Timma5	-66,658	6,243	-10,678	,000
Timma6	-50,147	6,243	-8,033	,000
Timma7	-29,678	6,243	-4,754	,000
Timma8	-1,487	6,243	-,238	,812
Timma9	22,148	6,243	3,548	,000
Timma10	22,404	6,243	3,589	,000
Timma11	22,240	6,243	3,563	,000
Timma12	16,482	6,243	2,640	,008
Timma13	8,382	6,243	1,343	,179
Timma15	-8,738	6,243	-1,400	,162
Timma16	-13,390	6,243	-2,145	,032
Timma17	-11,573	6,243	-1,854	,064
Timma18	-1,073	6,243	-,172	,864
Timma19	4,532	6,243	,726	,468
Timma20	-1,985	6,243	-,318	,751
Timma21	-7,237	6,243	-1,159	,246
Timma22	-9,605	6,243	-1,539	,124
Timma23	-12,043	6,243	-1,929	,054
Timma24	-30,014	6,243	-4,808	,000

a. Dependent Variable: Timpris

Parameterskattningar för dummyvariablerna med timma 14 som referensvariabel

Tabell 1.2. *Dickey-Fuller test för timpris efter säsongsdifferentiering*

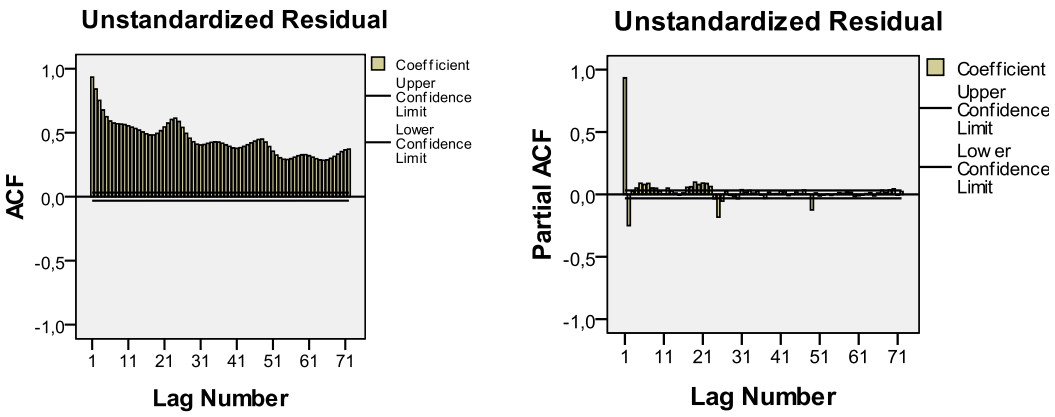
Coefficients[a,b]

Model		Unstandardized Coefficients		Sig.
		B	Std. Error	
1	LAGS(RES_1,24)	-,385	,012	,000

a. Dependent Variable: SDIFF(RES_1,1,24)

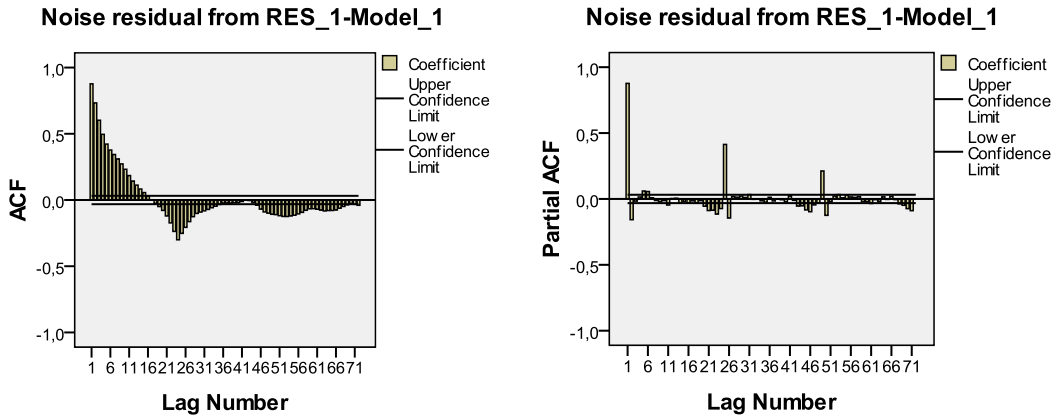
b. Linear Regression through the Origin

Figur 1.1



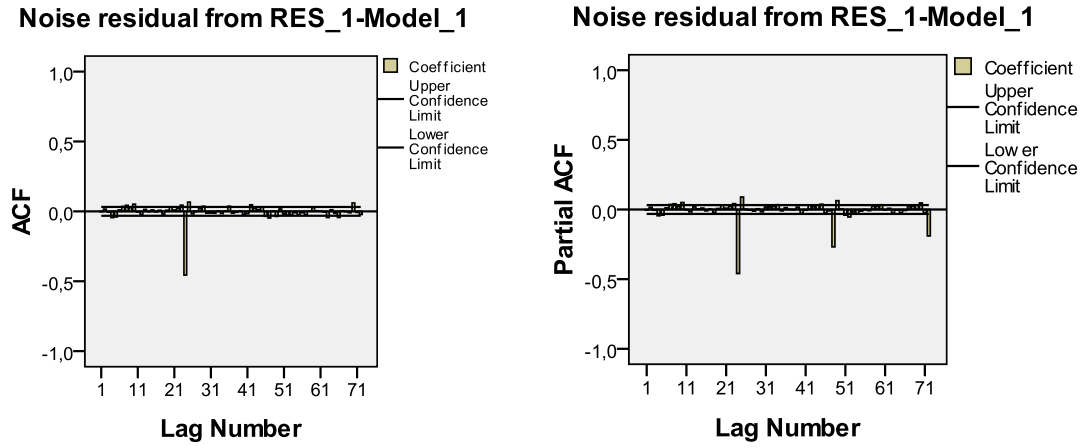
Korrelogram för ACF och PACF över residualerna efter att dummyvariablerna skattats

Figur 1.2



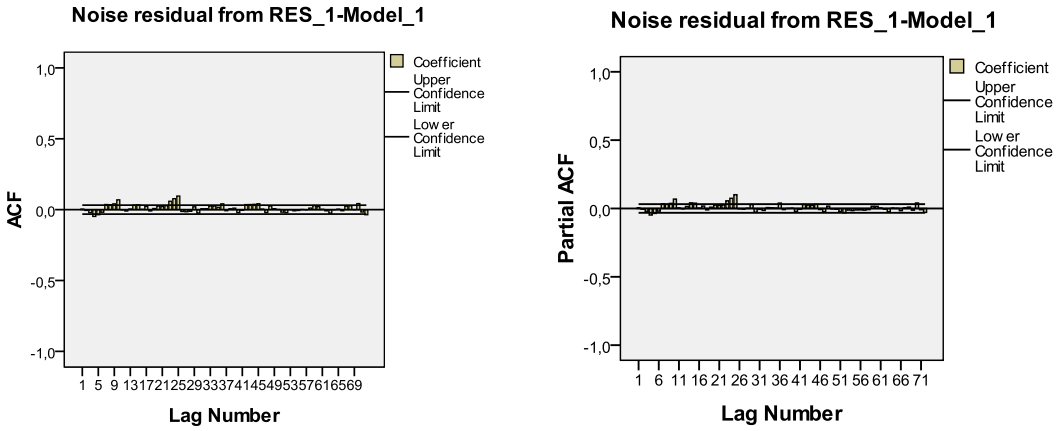
Korrelogram för ACF och PACF efter insättandet av en SARIMA $(0,0,0)(0,1,0)_{s=24}$

Figur 1.3



Korrelogram för ACF och PACF efter insättandet av en SARIMA $(2,0,0)(0,1,0)_{s=24}$

Figur 1.4



Korrelogram för ACF och PACF efter insättandet av en SARIMA $(2,0,0)(0,1,1)_{s=24}$

Tabell 1.3. *Modellsummering SARIMA* $(2,0,0)(0,1,1)_{s=24}$

Model Statistics

Model		Model Fit statistics		Ljung-Box Q(18)	
	Stationary R-squared	AIC	Normalized BIC	Statistics	Sig.
SARIMA(2,0,0)(0,1,1)	,860	34630,826	5,863	65,673	,000

ARIMA Model Parameters

			Estimate	SE	t	Sig.
SARIMA(2,0,0)(0,1,1)	Constant		-,478	,353	-1,354	,176
	AR	Lag 1	1,099	,016	70,834	,000
		Lag 2	-,204	,016	-13,154	,000
	Seasonal Difference		1			
	MA, Seasonal	Lag 1	,880	,008	112,140	,000

Tabell 1.4. *Modellmått SARIMA* $(2,0,0)(0,1,1)_{s=24}$

$SARIMA\,(p,d,q)(P,D,Q)_{s=12}$	*Stat.* R^2	*Normalized BIC*	*AIC*	*Ljung-Box*
$SARIMA\,(0,0,0)(0,1,0)_{s=12}$	$-2{,}2\cdot10^{-16}$	7,82	42452	10603,3
$SARIMA\,(2,0,0)(0,1,0)_{s=12}$	0,775	6,33	36509	46,756
$SARIMA\,(2,0,0,)(0,1,1)_{s=12}$	0,860	5,863	34631	65,673

Tabell 1.5. *Prognosvärden med 95%igt prediktionsintervall samt de faktiska värdena under den 17e juni 2009*

Timma	Prognosvärden	Övre Pi	Nedre Pi	Faktiska värden
1	356,7027728	393,089	320,316	330,98
2	337,3630186	391,467	283,259	296,56
3	317,5071446	382,833	252,181	268,33
4	300,3032509	373,083	227,523	270,17
5	288,4940198	366,406	210,582	215,66
6	317,063326	398,596	235,53	313,28
7	352,0021749	436,134	267,871	388,64
8	383,4714444	469,488	297,455	413,62
9	408,5144156	495,909	321,12	430,12
10	423,7073265	512,114	335,3	445,87
11	434,8212788	523,975	345,668	457,38
12	443,1854622	532,892	353,479	469,43
13	431,1448418	521,26	341,029	493
14	419,6456754	510,065	329,226	494,63
15	408,0758698	498,721	317,431	492,02
16	395,7681488	486,581	304,956	460,75
17	387,9418954	478,879	297,005	445,22
18	390,7290982	481,759	299,699	435,23
19	394,5182904	485,617	303,419	426,43
20	395,4226141	486,573	304,272	419,27
21	389,4809737	480,67	298,292	412,97
22	390,6676458	481,885	299,45	399,61
23	400,5641268	491,803	309,325	407
24	372,1821978	463,437	280,928	384,08

Bilaga 2. Modellbyggnadsprocessen för dagsmedelpris

Tabell 2.1. Modellsummering $LogX_t = \alpha + \gamma_i I_i$

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,113[a]	,013	,012	,54124

a. Predictors: (Constant), Sun, Sat, Fri, Thu, Wed, Tue

b. Dependent Variable: Logdagsmedel

Coefficients[a]

Model		Unstandardized Coefficients		t	Sig.
		B	Std. Error		
1	(Constant)	5,461	,020	267,508	,000
	Tue	,006	,029	,216	,829
	Wed	,002	,029	,057	,955
	Thu	-,010	,029	-,348	,728
	Fri	-,033	,029	-1,159	,246
	Sat	-,115	,029	-3,979	,000
	Sun	-,161	,029	-5,572	,000

a. Dependent Variable: Logdagsmedel

Parameterskattningar för dummyvariablerna med måndag som referensvariabel

Tabell 2.2. Dickey-Fuller test *för logaritmerat dagsmedelpris efter säsongsdifferentiering*

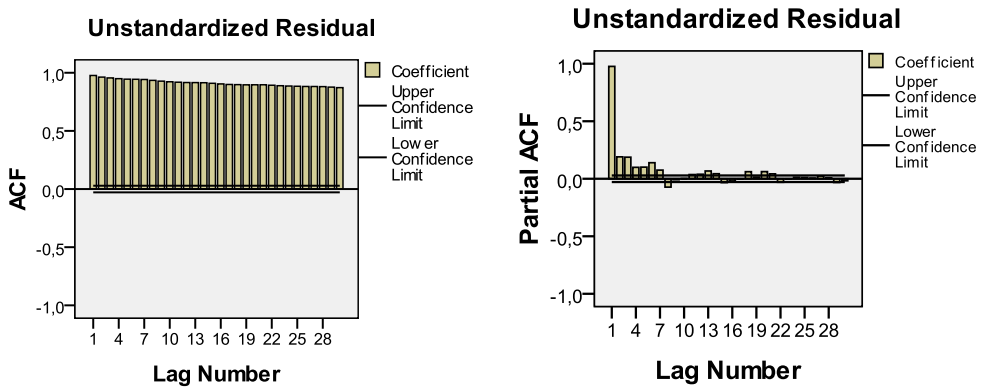
Coefficients[a,b]

Model		Unstandardized Coefficients		
		B	Std. Error	Sig.
1	LAGS(RES_1,7)	-,056	,005	,000

a. Dependent Variable: SDIFF(RES_1,1,7)

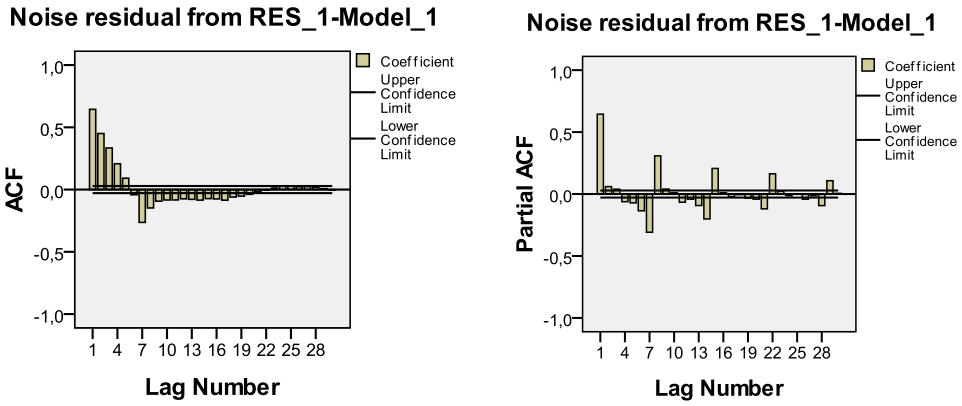
b. Linear Regression through the Origin

Figur 2.1



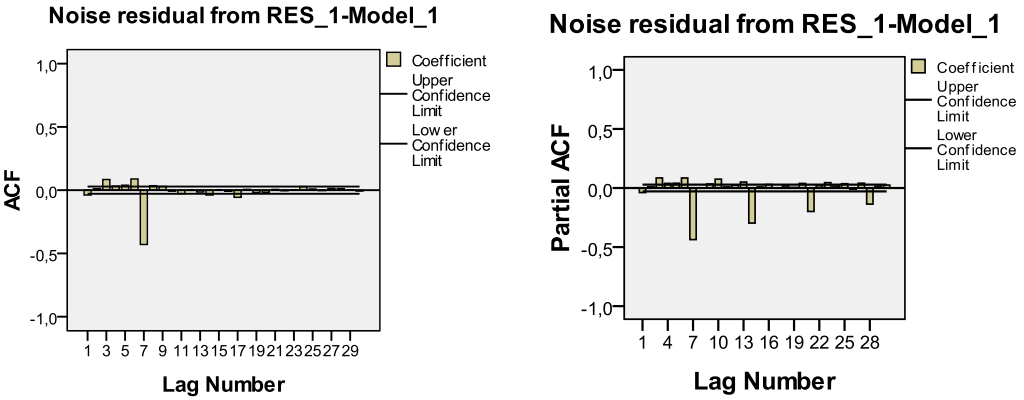
Korrelogram för ACF och PACF över residualerna efter att datamaterialet logaritmerats och dummyvariablerna skattats

Figur 2.2



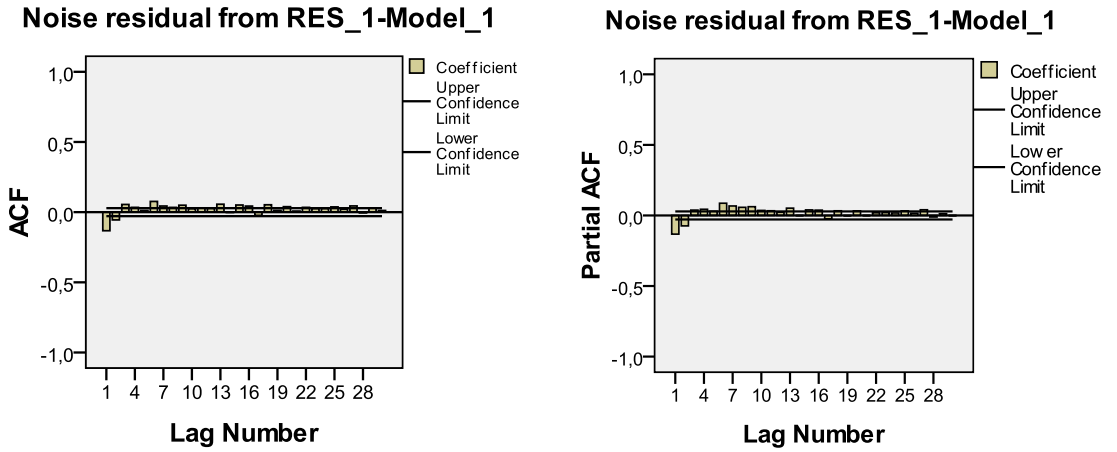
Korrelogram för ACF och PACF efter insättandet av en SARIMA $(0,0,0)(0,1,0)_{s=7}$

Figur 2.3



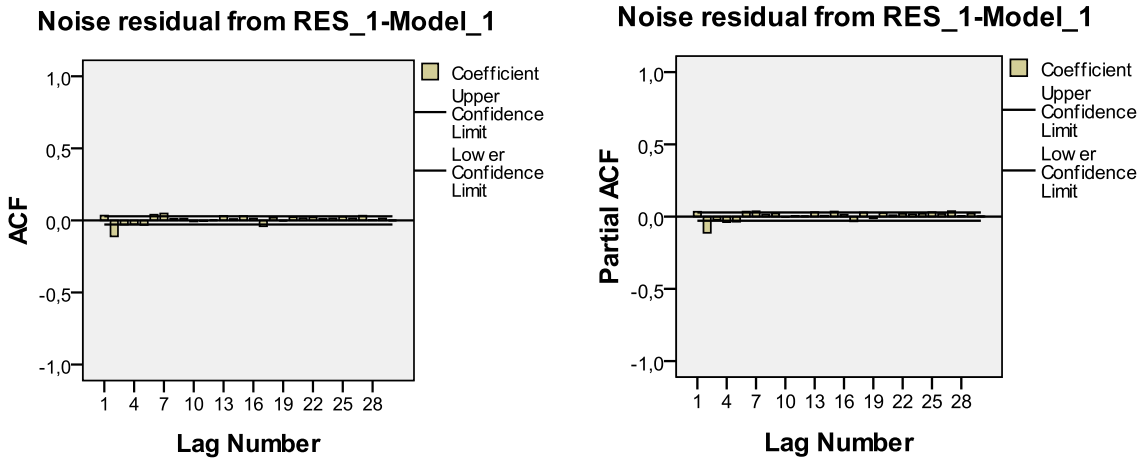
Korrelogram för ACF och PACF efter insättandet av en SARIMA $(1,0,0)(0,1,0)_{s=7}$

Figur 2.4



Korrelogram för ACF och PACF efter insättandet av en SARIMA $(1,0,0)(0,1,1)_{s=7}$

Figur 2.5



Korrelogram för ACF och PACF efter insättandet av en SARIMA $(1,0,1)(0,1,1)_{s=7}$

Tabell 2.3. Modellsummering $SARIMA(1,0,1)(0,1,1)_{s=7}$

Model Statistics

Model	Model Fit statistics			Ljung-Box Q(18)	
	Stationary R-squared	AIC	Normalized BIC	Statistics	Sig.
SARIMA(1,0,1)(0,1,1)	,629	-7612	-4,383	120,184	,000

			Estimate	SE	t	Sig.
SARIMA(1,0,1)(0,1,1)	Constant		,001	,002	,390	,697
	AR	Lag 1	,932	,006	149,684	,000
	MA	Lag 1	,275	,016	17,623	,000
	Seasonal Difference		1			
	MA, Seasonal	Lag 1	,874	,007	117,830	,000

ARIMA Model Parameters

Tabell 2.4. Modellmått $SARIMA(1,0,1)(0,1,1)_{s=7}$

$SARIMA(p,d,q)(P,D,Q)_{s=12}$	*Stat.* R^2	*Normalized BIC*	*AIC*	*Ljung-Box*
$SARIMA(0,0,0)(0,1,0)_{s=12}$	$-2{,}69\cdot10^{-15}$	-3,397	-2754	4609,69
$SARIMA(1,0,0)(0,1,0)_{s=7}$	0,415	-3,937	-5387	1039,67
$SARIMA(1,0,0)(0,1,1)_{s=7}$	0,614	-4,346	-7426	242,09
$SARIMA(1,0,1)(0,1,1)_{s=7}$	0,629	-4,383	-7612	120,18

Tabell 2.5. *Prognosvärden tre veckor framåt med 95%igt prediktionsintervall samt de faktiska värdena från och med den 17e juni 2009*

Dag	Prognosvärden	Övre Pi	Nedre Pi	Faktiska värden
Mån				402,03
Tis	411,0849976	511,322241	330,4977991	398,7604167
Ons	408,666737	530,542464	314,78819	362,6029167
Tor	407,6871129	546,5359021	304,1131998	351,37
Fre	393,5861592	541,1518915	286,2598601	295,24125
Lör	361,7668704	507,7047105	257,8041015	273,4379167
Sön	347,0261023	495,2684803	243,1795356	409,0208333
Mån	400,4544391	579,5198988	276,7182941	395,965
Tis	409,5053631	604,1847327	277,552471	406,0045833
Ons	407,2389036	609,2934456	272,1626866	398,1408333
Tor	406,3845993	615,2323492	268,4057961	399,10625
Fre	392,4071701	600,0422084	256,620926	350,7670833
Lör	360,7914173	556,4069773	233,9716607	370,36
Sön	346,1942383	537,7533512	222,8725314	434,4195833
Mån	399,5744077	624,4679455	255,673183	447,6654167
Tis	408,666737	645,2901109	258,8115006	453,9933333
Ons	406,5065329	646,5819826	255,5453784	456,9854167
Tor	405,6943323	649,4332114	253,4584743	449,48
Fre	391,8581845	630,6178165	243,4958745	410,47875
Lör	360,3226932	582,5412509	222,8725314	388,03
Sön	345,7790544	561,2688364	213,0229699	484,3504167

Bilaga 3. Modellbyggnadsprocessen för månadsmedelpriser

Tabell 3.1. Modellsummering $LogX_t = \alpha + \gamma_i I_i$

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,159[a]	,025	,007	,51809

a. Predictors: (Constant), Kvartal4, Kvartal3, Kvartal2

Coefficients[a]

Model		Unstandardized Coefficients		t	Sig.
		B	Std. Error		
1	(Constant)	5,474	,080	68,473	,000
	Kvartal2	-,138	,114	-1,212	,227
	Kvartal3	-,114	,115	-,993	,322
	Kvartal4	,064	,115	,554	,580

a. Dependent Variable: LogMånadsmedel

Parameterskattningar för dummyvariablerna med kvartal 1 som referensvariabel

Tabell 3.2. *Dickey-Fuller test för logaritmerat månadsmedelpris efter säsongsdifferentiering*

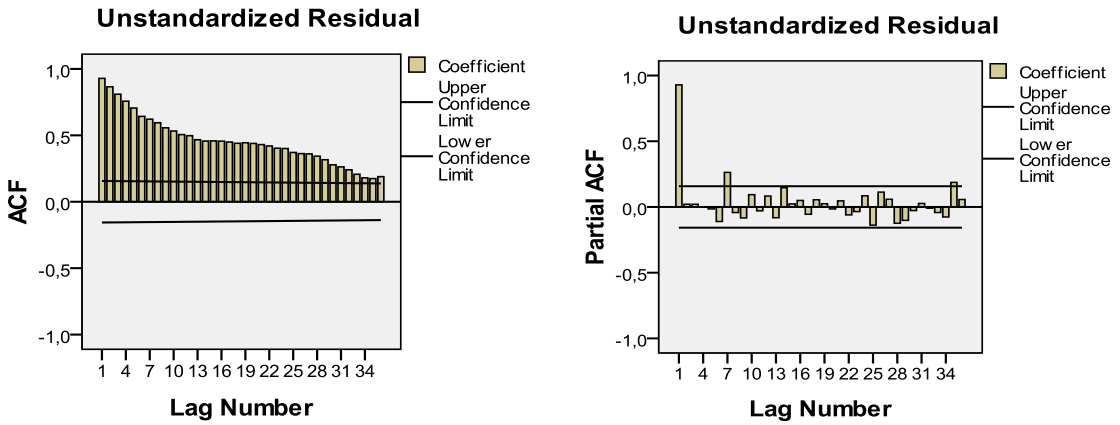
Coefficients[a,b]

Model		Unstandardized Coefficients		Sig.
		B	Std. Error	
1	LAGS(RES_1,12)	-,392	,075	,000

a. Dependent Variable: SDIFF(RES_1,1,12)

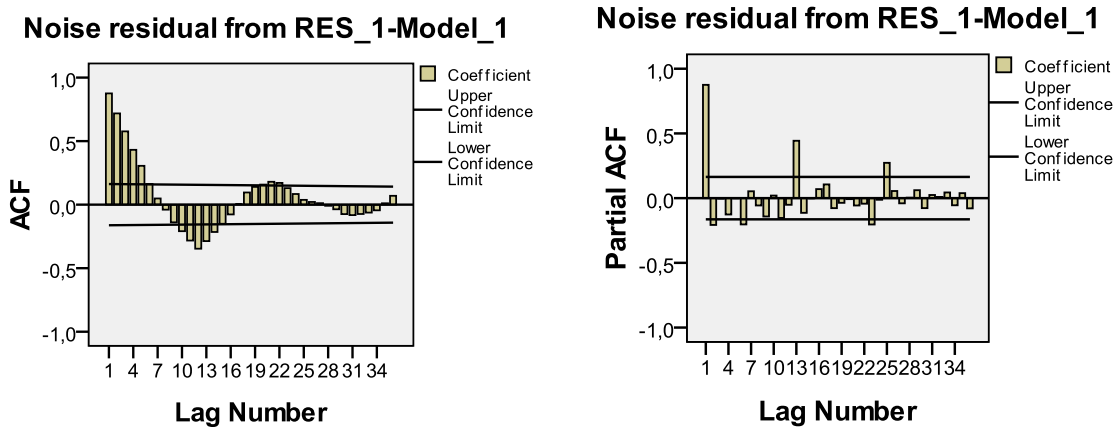
b. Linear Regression through the Origin

Figur 3.1



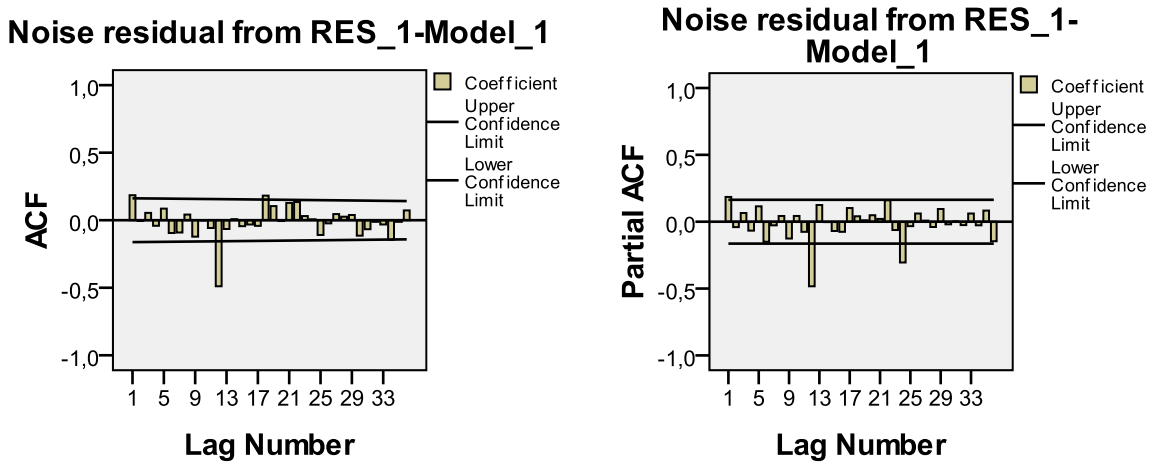
Korrelogram för ACF och PACF över residualerna efter att datamaterialet logaritmerats och dummyvariablerna skattats

Figur 3.2



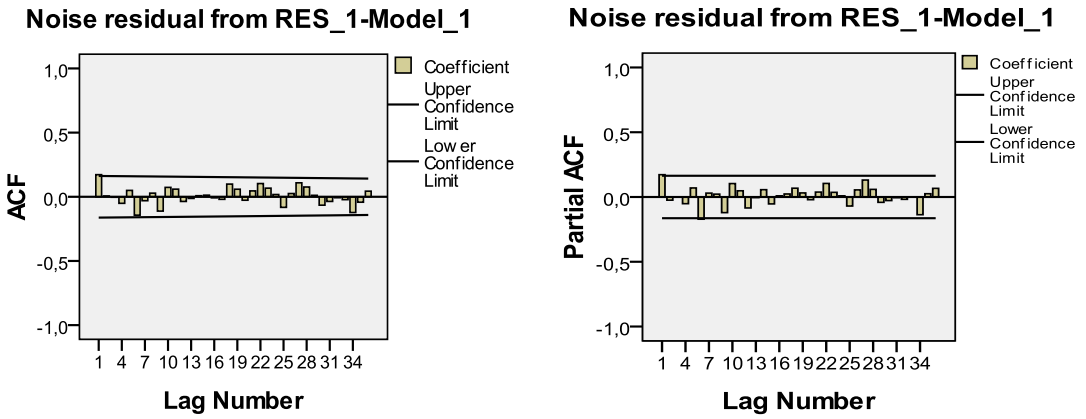
Korrelogram för ACF och PACF efter insättandet av en SARIMA $(0,0,0)(0,1,0)_{s=12}$

Figur 3.3



Korrelogram för ACF och PACF efter insättandet av en SARIMA $(1,0,0)(0,1,0)_{s=12}$

Figur 3.4



Korrelogram för ACF och PACF efter insättandet av en SARIMA $(1,0,0)(0,1,1)_{s=12}$

Tabell 3.3. Modellsummering $SARIMA(1,0,0)(0,1,1)_{s=12}$

Model Statistics

Model		Model Fit statistics		Ljung-Box Q(18)	
	Stationary R-squared	AIC	Normalized BIC	Statistics	Sig.
SARIMA (1,0,0)(0,1,1)	,858	-85,019	-3,316	14,300	,576

ARIMA Model Parameters

			Estimate	SE	t	Sig.
SARIMA (1,0,0)(0,1,1)	Constant		,082	,032	2,521	,013
	AR	Lag 1	,910	,037	24,658	,000
	Seasonal Difference		1			
	MA, Seasonal	Lag 1	,925	,169	5,458	,000

Tabell 3.4. Modellmått $SARIMA(1,0,0)(0,1,1)_{s=12}$

SARIMA $(p,d,q)(P,D,Q)_{s=12}$	***Stat.*** R^2	***Normalized BIC***	***AIC***	***Ljung-Box***	***Sig. (L.B.)***
$SARIMA(0,0,0)(0,1,0)_{s=12}$	$-3{,}405\cdot10^{-16}$	-1,443	203,789	365,563	
$SARIMA(1,0,0)(0,1,0)_{s=12}$	0,766	-2,857	-9,527	59,512	
$SARIMA(1,0,0)(0,1,1)_{s=12}$	0,858	-3,316	-85,019	14,3	0,576

Tabell 3.5. Prognostisering 31 månader framåt från och med juni 2009 *med 95%igt prediktionsintervall*:

Månad		Prognosvärden	Övre Pi	Nedre Pi
	6	385,5419474	537,2749636	276,6601894
	7	340,8764176	533,8420454	217,6590832
	8	418,9828549	708,9039182	247,6310656
	9	441,1169354	789,4764466	246,4699304
	10	451,4253403	843,0868689	241,7127409
	11	474,8409947	916,6172578	245,9848626
	12	482,8132824	956,5662027	243,6931861
	1	474,0534125	958,7976005	234,3861573
	2	448,370545	922,2626013	217,9814571
	3	415,2289272	865,8225674	199,1344052
	4	404,1717858	852,1648551	191,6958497
	5	377,446939	803,020305	177,4111642
	6	415,5570877	895,1130293	192,9247082
	7	367,6384565	799,8706283	168,974869
	8	452,1301134	991,7191973	206,126488
	9	476,2533688	1051,496851	215,7089402
	10	487,6071202	1082,241898	219,6927545
	11	513,1150043	1143,718414	230,2026482
	12	521,9229648	1167,359921	233,3501231
	1	512,6329028	1149,750223	228,5648551
	2	485,010238	1090,967286	215,6183615
	3	449,2861546	1012,593358	199,3475931
	4	437,4401953	987,444211	193,7870741
	5	408,6095277	923,5177317	180,7870827
	6	449,9650891	1020,216359	198,4565134
	7	398,1544519	905,0588779	175,1565245
	8	489,7475234	1115,524526	215,0111742
	9	515,9603905	1177,1011	226,1616479
	10	528,3399916	1206,815076	231,3056509
	11	556,0509907	1271,306003	243,2087189
	12	565,6638516	1294,189617	247,2425899