



Kandidatuppsats

Statistiska institutionen

Bachelor thesis, Department of Statistics

Nr 2023:1

Multivariat och tidsserieanalys av elspotpriser i Sverige

***Multivariate and time series analysis of
Electricity prices in Sweden***

Isak Mars och Rashiq Al Tariq

Självständigt arbete 15 högskolepoäng inom Statistik III, VT2022

Handledare: Per Gösta Andersson

Sammanfattning

På grund av de stigande och oförutsägbara elpriserna som har uppstått under de senaste åren har intresset vuxit angående frågan om elpriser och om deras förändring går att tillförlitligt modellera. I Sverige finns det fyra elområden som är fokuset av denna analys. Elpriserna i de fyra områdena bestäms en gång i timmen dygnet runt. Medelvärden av dessa så kallade spotpriser har beräknats, vilket ger en observation per dag. Analysen i denna rapport har utförts på dessa medelvärden.

Syftet med denna rapport är att hitta underliggande faktorer och mönster kring prissättningen av el som kan observeras för att utföra den tidigare nämnda modelleringen. Slutsatserna dragna i denna uppsats har gjorts huvudsakligen utifrån genomförandet av en utforskande faktoranalys samt ARIMA och GARCH tidsseriemodellering.

Den utförda faktoranalysen identifierade en lämplig tvåfaktorstruktur till data. Där den första faktorn representerar allmän elförbrukning och den andra representerar den geografiska skillnaden i elproduktionskapacitet relativt till efterfrågan av el.

Tidsserieanalysen inleds med passning av ARIMA- och sedan GARCH-modeller. Då elpriserna uppvisar en stor mängd volatilitet kan slutsatsen dras genom jämförelse av AIC-värden att GARCH-modeller är signifikant bättre på att predicera elpriser än sina ARIMA motparter. De uppskattade GARCH-modellerna används sedan för att predicera elspotpriser med hjälp av tidigare observationer. Resultaten visade att GARCH-modellerna lyckades fånga tidsseriernas generella mönster för hela tidsspannet.

Abstract

Due to the rising and unpredictable price of electricity that has become more common over the last few years, interest in the questions concerning the price of electricity and whether it is possible to reliably model the changes in them has grown. In Sweden, there are four power grid areas. These four areas will be the focal point of this report. The electricity prices in these areas are decided once an hour every day, these are called spot prices. The average of these so-called spot prices have been calculated, resulting in one observation per day. All the following analysis has been performed using these averages.

This report aims to find underlying factors and patterns that can be observed to perform the previously mentioned modeling of spot prices. The conclusions drawn in this study have been made mainly through the use of exploratory factor analysis, as well as ARIMA and GARCH time series modeling.

The performed factor analysis identified a suitable two-factor structure to data. Where the first factor represents general electricity consumption, and the second represents the geographical difference in electricity production capacity relative to the demand of electricity.

The time series analysis starts with fitting ARIMA models for each power grid area and is followed by fitting GARCH models in the same fashion. Through a comparison of AIC values between the different types of models, the conclusion can be drawn that GARCH models should be significantly better at predicting electricity prices than their ARIMA counterparts. The estimated GARCH models are then used to predict the electricity spot prices with the help of earlier observations. The results show that the GARCH models are able to capture the general pattern of the entire time series.

Förord

Vi skulle vilja tacka Per Gösta Andersson för hans kontinuerliga vägledning och råd, Nord Pool för tillhandahållning av datamaterial, samt Niklas Åström och Mats Grönmark för hjälp med korrekturläsning.

Innehållsförteckning

1	Inledning	6
2	Bakgrund.....	6
3	Syfte och frågeställning	7
4	Tidigare forskning	8
5	Metod	9
5.1	Data	9
5.2	Multivariat analys	9
5.2.1	Utforskande faktoranalys.....	9
5.2.2	Faktoranalysmetoder.....	11
5.2.3	Mått för goodness-of-fit.....	12
5.2.4	Utvärdering av data – Lämpligheten av utforskande faktoranalys	12
5.3	Tidsserieanalys.....	14
5.3.1	ARIMA	14
5.3.2	ARCH och GARCH	16
6	Empirisk undersökning och resultat	17
6.1	Faktoranalys	17
6.2	Tidsserieanalys.....	21
6.2.1	ARIMA	21
6.2.2	GARCH	23
6.2.3	Predicering.....	26
7	Diskussion och slutsats	30
8	Framtida forskning.....	31
	Litteraturförteckning.....	32
	Bilaga A: Fullständiga tidsserier	34

1 Inledning

Priset av el för hushåll, likt prisnivån i samhället över lag, har ökat markant de senaste månaderna. Under den kalla vintertiden förväntas behovet av energi att öka för att förse företag och hushåll med värme och ljus, men den aktuella prisökningen beror på mer än bara den förväntade ökningen i efterfrågan.

Denna prisutveckling har uppstått på grund av ett antal händelser vars efterföljder har haft en global påverkan. Den främsta orsaken är Rysslands fullskaliga invasion av Ukraina som skedde i februari 2022 (Damberg, 2023). Den ryska invasionen har haft en påtaglig effekt på energimarknaden genom ökad oro som reflekteras i ökade energipriser, men också genom Rysslands avstängning av export av naturgas. Beroendet av rysk energi har visat sig vara en stor brist hos Europas länder i allmänhet. Ytterligare en orsak till den rådande situationen kring elpriser är den globala ekonomin. Efter den ekonomiska påfrestningen som tillfogats många länder av covid-19 pandemin har den ekonomiska återhämtningen lett till högre priser på fossila bränslen, vilket i sin tur har lett till högre energipriser.

Med den rådande situationen skulle det vara intressant att undersöka huruvida det är möjligt att identifiera eventuella mönster kring prissättningen av el i Sverige, eller fastställa om utvecklingen av energipriser är alltför volatil för att kunna göra någon sådan bedömning.

Med hjälp av ett antal metoder kommer denna rapport undersöka huruvida spotpriset för el i Sverige går att effektivt beskriva med hjälp av statistiska tidsseriemodeller, samt om det finns några underliggande mönster i elpris mellan de fyra elområdena i Sverige.

2 Bakgrund

År 1995 etablerades en gemensam norsk-svensk elbörs vid namn Nord Pool ASA (Nord Pool, n.d.). Denna etablering skedde till följd av Norges och Sveriges avreglering av elhandel som resulterade i förenandet av dessa två länders elmarknader. Enligt den svenska regeringen var syftet med avregleringen att öka konkurrensen inom elmarknaden och sänka priser för konsumenter (Riksdagen, n.d.). Under de nästkommande åren skulle den nordiska elmarknaden fullständigt integreras då Danmark och Finland också anslöt sig till den nya marknadsplatsen. Numera är Nord Pool en internationell marknadsplats för handel av el för ett antal europeiska länder.

Prissättningen av el på elbörsen sker genom ett antal processer. Huvudsakligen bestäms elpriset för nästa dag genom auktionshandel där producenter och köpare lämnar bud för de priser de vill sälja och köpa för. Med hjälp av dessa bud bestämmer Nord Pool ett så kallat Spotpris, som konstant förändras på grund av tillgång och efterfrågan hos producenter och köpare. Denna tillgång och efterfrågan, och vad som påverkar dem, är de externa faktorer som kommer att vara viktiga för oss. Spotpriset är alltså det pris som elbolag köper in el för och ligger därmed till grund för det pris som konsumenter betalar för sin el. Med priserna som sätts av Nord Pool sker handeln med el i Sverige som beskrivs ovan huvudsakligen på Stockholmsbörsen NASDAQ.

I Sverige finns det fyra elområden:

- SE1 (Luleå)
- SE2 (Sundsvall)
- SE3 (Stockholm)
- SE4 (Malmö)

Enligt Svenska kraftnät är uppdelningen ”ett sätt att hantera fysiska begränsningar, så kallade flaskhalsar i transmissionsnätet.” (Svenska Kraftnät, n.d.). Flaskhalsar i transmissionsnätet är områden där nätet inte fysiskt klarar av att överföra så mycket el som det efterfrågas. Indelningen i elområden baseras på var det finns överföringsbegränsningar i elnätet som ofta riskerar att överbelastas. Enligt energimarknadsinspektionen kännetecknas de två första områdena av att det finns ett överskott av elproduktion relativt till efterfrågan, medan de resterande två områdena i stället upplever ett underskott av elproduktion relativt till efterfrågan (Energimarknadsinspektionen, 2021). Svenska kraftnät agerar som systemansvarig för det svenska överföringssystemet av el. De arbetar med att granska uppdelningen av elområdena för att säkerställa korrekt prissättning för alla aktörer på marknaden oavsett vart på nätet man befinner sig, effektiv hantering av överbelastningar i överföringssystemet, samt förutsättningar för säker och kostnadseffektiv drift och utveckling av kraftsystemet.

3 Syfte och frågeställning

Denna studie ämnar undersöka utvecklingen av elpriser i syfte att identifiera underliggande faktorer som påverkar prissättningen av el samt skapa en modell för att predicera framtida elpriser. Detta kommer att göras genom en multivariat analys och tidsserieanalys av svenska elprisdata. Rapportens frågeställningar är de följande:

Går det att identifiera ett antal faktorer som förklarar variationen av elpriserna i Sverige?

Är det möjligt att uppskatta en tillförlitlig modell för att predicera framtida elpriser?

4 Tidigare forskning

En tidigare kandidatuppsats av Fagerholm (2019) har varit en mycket givande referens för att bilda sammanhang kring hur elnätet fungerar och hur modelleringen av de fysiska elnäten skulle se ut. I studien jämförs ARIMA-modeller med så kallade LSTM-modeller för att predicera den dagliga elkonsumtionen. Resultatet visar att LSTM-modeller presterade bättre jämfört med ARIMA-modeller.

En mer avancerad studie på detta ämne har utförts av Andersson & Lillestøl (2010). I studien användes den timvisa prissättningen på Nord Pool som en variabel i 24-dimensioner, en för varje timme av dygnet. På det insamlade datamaterialet utfördes sedan en Multivariate Analysis of Variance (MANOVA). Efter den initiala analysen har genomförts passas en Auto Regressive Integrated Moving Average (ARIMA)-modell till resultatet av MANOVA-analysen. En “true multivariate state space scheme” passas sen. Resultatet var att ett antal kategorier kunde utvinnas från datamaterialet och som sedan kunde, tillsammans med hjälp av en dynamisk multivariat linjär regressionsmodell med tidsvarierande parametrar, användas för predicering av spotpriset. En kommentar i texten erkänner dock att modellen som mest kan användas i handelssammanhang för att inte gå med förlust, hellre än att erhålla vinster.

En annan uppsats som behandlar ämnet av elpriser är Gustavsson (2009). Författaren ämnar i denna uppsats att försöka predicera framtida elpriser på den nordiska elbörsen med ARIMA-modeller. Enligt resultatet visar det sig att modellernas prediktioner ligger nära de sanna värdena när det kommer till timpriserna och dagspriserna, dock lyckas inte modellerna fånga extremvärdena i data.

Utöver dessa specifika exempel är modellering av marknadsdata ett av de största områdena inom ekonometri och tidsserieanalys. Att med reliabilitet modellera nästa händelse i råvaru- och finansmarknader har länge varit en önskedröm som många försökt uppfylla utan att helt lyckas.

5 Metod

5.1 Data

Data för analysen i denna rapport har erhållits från Nord Pool. Datamaterialet omfattar det genomsnittliga spotpriset per dag mellan 1 januari 2019 och 31 mars 2023. Detta spann har valts specifikt för att kunna ta hänsyn till den globala COVID-pandemin och Rysslands fullskaliga invasion av Ukraina på spotpriset av el, datamaterialet täcker således elpriser innan och efter dessa händelser. Analys av materialet har skett löpande sedan 31 mars 2023 och på grund av detta har beslutet tagits att begränsa insamlingen efter det datumet så att resultaten inte förändras under arbetets gång.

I datamaterialet syns spotpriset för de fyra svenska elområdena i separata kolumner med priset i öre (SEK) per kilowattimme. Detta har sammanställts från separata datafiler för varje år och elområde.

Spotpriset sätts varje timme under hela dygnet, men att titta på alla de datapunkterna anses vara onödigt för syftet av detta arbete. Valet har därför gjorts att titta på snittpriset för varje dygn och i stället jämföra dessa för de fyra elområdena i Sverige. Snittpriset per dygn har redan beräknats i datafilerna försedda av Nord Pool vilket minskar antalet möjliga felkällor då datamaterialet inte har behövt manipuleras för att utföra analysen.

Varje rad innehåller därför medelvärdet av spotpriset under 24 timmar för varje elområde och sitter ihop med tillhörande datum.

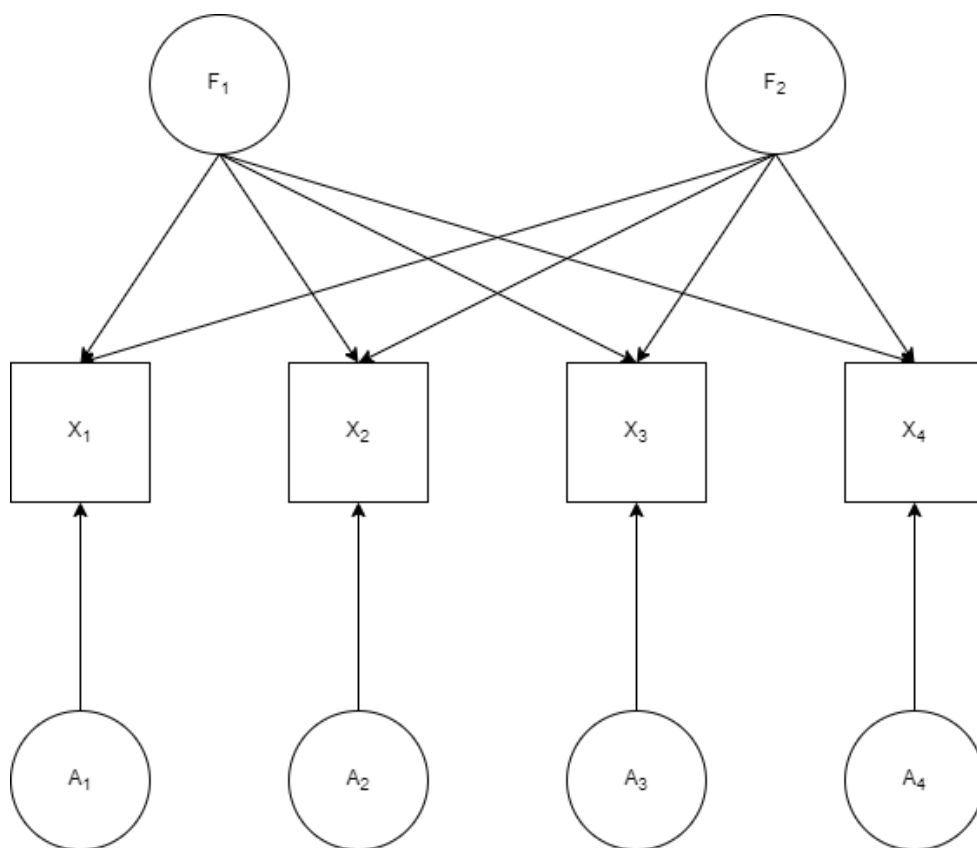
5.2 Multivariat analys

För den multivariata analysen kommer utforskande faktoranalys att användas.

5.2.1 Utforskande faktoranalys

Syftet med utforskande faktoranalys är att identifiera underliggande faktorer som förklarar gemenskapen mellan korrelerade indikatorer i ett dataset. I detta fall skall resultatet från den utförda faktoranalysen avslöja om det verkar finnas underliggande faktorer som förklarar sambandet mellan elpriserna i Sveriges olika elområden. Denna metod lämpar sig bäst till denna analys eftersom någon faktorstruktur inte antas i data, det anges alltså inga "a priori" begränsningar på parametrarna.

Några av de viktigaste koncepten bakom resultatet av en utforskande faktoranalys är indikatorernas laddningar och kommunalitet med varje faktor, indikatorernas unika varians, samt korrelationen mellan faktorerna. En faktormodell kan visualiseras som i Figur 5.2.1



Figur 5.2.1 Exempel: Struktur av faktormodell

Där F_1 och F_2 är de underliggande faktorerna, X_i är indikatorer och A_i är deras unika varians. Målet med faktoranalys är alltså att erhålla en liknande struktur till figuren ovan för det datamaterial som analyseras. Laddningarna representerar korrelationen mellan en indikator och en faktor, kommunaliteten representerar den delade variansen mellan en indikator och en faktor, och den unika variansen representerar variationen i indikatorerna som inte förklaras av faktorerna.

Generellt sätt kan en indicators värde representeras som en linjär kombination av de underliggande faktorerna och indikatorns unika varians.

$$X_i = aF_1 + bF_2 + A_i \quad (5.2.1)$$

Koefficienterna a och b i ekvation 5.2.1 är indikatorernas mönsterladdningar.

5.2.2 Faktoranalysmetoder

Vid tillämpningen av utforskande faktoranalys finns det ett antal olika metoder för att erhålla en faktorstruktur. De två metoder som kommer att utnyttjas i detta arbete är principal component factoring (PCF) och principal axis factoring (PAF).

PCF

Det första steget vid både PCF och PAF är att estimerar kommunaliteterna. Vid PCF antas dessa vara lika med 1. Utgångspunkten för PCF är alltså att det inte finns någon unik faktor och att antalet komponenter/faktorer är lika med antalet indikatorer. Sedan utsätts korrelationsmatrisen med de antagna kommunaliteterna på diagonalen för en principalkomponent analys (PCA). PCA är en dimensionsreducerande metod som ämnar till att summera data med många variabler till ett mindre antal "indexvariabler" som representerar variationen i data. Efter att PCA har utförts utvärderas det hur många principalkomponenter (faktorer) som skall behållas. De komponenter som behålls antas vara variablernas gemensamma faktorer, medan de resterande komponenterna sätts ihop till en enda komponent som representerar en "unik" faktor, den delade variansen mellan indikatorerna och denna unika faktor är indikatorernas unika varians. Faktörlösningen som erhålls från PCF är alltså ekvivalent med PCA. I denna studie kommer PCF enbart att användas för att avgöra hur många faktorer som bör tas med i den slutgiltiga faktormodellen.

PAF

PAF är en metod för faktoranalys som föredras av de flesta forskare (Sharma, 1995, p. 108), detta eftersom till skillnad från PCF så antas det att det finns en implicit underliggande faktormodell till data. Utgångspunkten för PAF är alltså att indikatorerna i modellen består av en gemensam del och en unik del, där den gemensamma delen beror på en gemensam faktor och den unika delen beror på en unik faktor. PAF är i praktiken en iterativ PCF process som tillämpas för att estimerar kommunaliteterna, denna process fortsätter tills estimationerna av kommunaliteterna konvergerar. Denna process beskrivs enligt följande steg:

- I den första iterationen antas det att kommunaliteterna är 1. PCF utförs på korrelationsmatrisen och med laddningarna erhållna estimeras kommunaliteterna.
- Den maximala skillnaden mellan de estimerade kommunaliteterna beräknas. Eftersom kommunaliteterna är lika med 1 för denna första iteration är den skillnaden lika med
$$1 - (\text{estimerad kommunalitet})$$
- Om den maximala skillnaden är större än något förbestämt konvergenskriterium, ersätts korrelationsmatrisens initiala kommunaliteter med de nya estimerade kommunaliteterna. Sedan, med de nya kommunaliteterna, repeteras det första och andra steget i processen tills kommunaliteterna konvergerar.

Efter att en PCF har utförts för att avgöra hur många faktorer som bör användas, kommer en PAF att utföras för att estimerar den slutliga faktormodellen.

5.2.3 Mått för goodness-of-fit

För att evaluera hur bra de givna faktormodellerna passar data kommer två mått att användas, residualanalys och RMSR.

Residualanalys

Efter att en given faktorstruktur har estimerats är det möjligt att utvärdera residual-korrelationsmatrisen. På diagonalen av denna matris finns indikatorernas unika varians medan de icke-diagonala elementen innehåller skillnaden mellan de observerade korrelationerna mellan indikatorerna och de reproducerade korrelationerna som förklaras av faktormodellen. För en god faktorlösning önskas små residualer.

RMSR

Root Mean Square Residual (RMSR) är en sammanfattning av residual-korrelationsmatrisen. Ekvationen för RMSR ges i 5.2.3.

$$RMSR = \frac{\sqrt{\sum_{i=1}^p \sum_{j=1}^p res_{ij}^2}}{p(p-1)2} \quad (5.2.3)$$

där res är lika med korrelationen mellan den i :te och j :te indikatorn, och p är antalet indikatorer. För en god faktorlösning önskas det ett lågt värde på detta mått.

5.2.4 Utvärdering av data – Lämpligheten av utforskande faktoranalys

I analysen kommer de fyra områdena användas som indikatorer i modellen. För att säkerställa lämpligheten av utforskande faktoranalys ytterligare examineras korrelationsmatrisen för dessa indikatorer. Eftersom syftet med utforskande faktoranalys är att identifiera en underliggande struktur till data som förklarar korrelationen mellan olika indikatorer är det nödvändigt att det finns ett tydligt samband mellan dem, det är alltså önskvärt att det råder homogenitet mellan indikatorerna.

Tabell 5.2.1 Korrelationsmatris och Kaiser-Meyer-Olkin measure of overall sampling adequacy (KMO) för elprisdata.

	Område 1	Område 2	Område 3	Område 4
Område 1	1.000	0.979	0.674	0.566
Område 2	0.979	1.000	0.697	0.603
Område 3	0.674	0.697	1.000	0.949
Område 4	0.566	0.603	0.949	1.000
KMO			0.64	

Matrisen i tabell 5.2.3 visar att korrelationerna mellan samtliga indikatorer är positiva och större än 0.5. Område 1 är starkt korrelerad med område 2, medan område 3 är starkt korrelerad med område 4. Detta tyder på att det finns två snarlika par av indikatorer i data, och att alla indikatorer är korrelerade med varandra.

Ett formellt mått på homogeniteten är "Kaiser-Meyer-Olkin measure of overall sampling adequacy (KMO)". Ett värde på KMO över 0.8 anses vara bra medan ett värde över 0.6 anses acceptabelt (Sharma, 1995, p. 116).

KMO för elprisdata är 0.64 vilket anses vara acceptabelt. Detta resultat i samband med utvärderingen av korrelationsmatrisen bekräftar att utforskande faktoranalys är en lämplig metod för denna analys.

5.3 Tidsserieanalys

5.3.1 ARIMA

I rapporten kommer ARIMA användas först för att uppskatta modeller. Detta på grund av att mönster och annan information datamaterialet innehåller kan bli synliga genom ARIMA-modellering för att senare användas för att underlätta GARCH-modelleringen som kommer att utföras.

ARIMA är en förkortning som står för Autoregressive Integrated Moving Average och är en sammansatt modell som består av tre delar med sina respektive parametrar. Den vanligaste notationen för ARIMA-modeller ser ut som:

$$ARIMA(p, d, q) \quad (5.3.1)$$

Där de tre parametrarna p , d , och q hör till sin respektive del av ARIMA-förkortningen. Till exempel är en modell av formen ARIMA (2,0,0) ekvivalent med en modell AR (2). Eller en "Autoregressive model of order 2". ARIMA-modeller används ofta för att beskriva tidsserier där målet är att eliminera säsongvariation eller trend. (Montgomery, 2016, p. 327)

Storleken av de olika parametrarna beskriver egenskaperna av den linjära ekvationen som representerar modellen. Parametrarna är alltid icke-negativa. "p" beskriver tidsförskjutningen, eller graden av den autoregressiva modelldelen, "d" eller "differencing" beskriver hur många gånger datan bör differentieras för att uppnå kovariansstationaritet. Till sist "q" som beskriver graden av glidande-medelvärdenmodellen (Moving Average).

Den första delen, Autoregressiva modellen, eller AR definieras som i ekvation 5.3.2

$$X_t = \sum_{i=1}^p \varphi_i X_{t-i} + \varepsilon_t \quad (5.3.2)$$

Där $\varphi_1, \dots, \varphi_p$ representerar parametrarna av modellen och t är vitt brus. Det går att se från ekvation 5.3.1 att antalet förskjutna tidsvariabler beror på parametervärdet "p".

MA-delen av en ARIMA-modell beskrivs i ekvation 5.3.3.

$$X_t = \mu + \sum_{i=1}^q \theta_i \varepsilon_{t-i} + \varepsilon_t \quad (5.3.3)$$

Här kan det observeras att parametervärdet interagerar med ekvationen på samma sätt som i AR-delen. Skillnaden ligger i att regressionen utförs på vitt brus-termerna i stället för tidigare värden av X .

Slutligen "I"-termen med parameter "d". Hur differentieringen utförs för att uppnå stationaritet ges av Ekvation 5.3.4.

$$(1 - L)^d X_t \quad (5.3.4)$$

Säsongsvariation i ARIMA-modeller beskrivs något annorlunda. Grundformen för detta syns i ekvation 5.3.5

$$ARIMA(p, d, q)X(P, D, Q)_m \quad (5.3.5)$$

Där parametrarna P , D , och Q beskriver samma som p , d , och q fast för säsongsvariationen och " m " beskriver antalet perioder per säsong i data.

R-funktionen `Auto.arima` från `forecast`-paketet kommer att användas för att estimerar de mest optimala modellerna för datamaterialet per område. `Auto.arima` använder AIC, AICc, och BIC kriterier för att bestämma den optimala modellen.

"Akaike Information Criterion" (AIC) och "Schwarz Bayesian Information Criterion" (BIC) är en estimator av prediktionsfel och därav den relativa pålitligheten av en statistisk modell för en given datamängd. Det är viktigt vid användningen av AIC att modellerna som jämförs måste ha skattats för exakt samma datamaterial.

AIC och BIC definieras som:

$$AIC = \ln \left(\frac{\sum_{t=1}^T e_t^2}{T} \right) + \frac{2p}{T} \quad (5.3.6)$$

$$BIC = \ln \left(\frac{\sum_{t=1}^T e_t^2}{T} \right) + \frac{p \ln(T)}{T} \quad (5.3.7)$$

Där " e " representerar residualerna, T representerar hur många perioder som finns i data, t är den perioden som observeras, och p är antalet parametrar som uppskattas i modellen (Montgomery, 2016, p. 76).

“Corrected Akaike Information Criterion” (AICc) är en modifierad variant av AIC som tar hänsyn till små stickprov. Då det finns en stor mängd observationer för alla områden kommer detta kriterium vara mindre relevant för denna analys. AICc definieras som:

$$\text{AICc} = \ln\left(\frac{\sum_{t=1}^T e_t^2}{T}\right) + \frac{2T(p+1)}{T-p-2} \quad (5.3.8)$$

AIC, BIC, och AICc ska alla minimeras för att utvärdera huruvida bäst modell har uppskattats.

Något viktigt att hålla i åtanke när dessa kriterier används för att skatta modellsignifikans är att det inte är absolutvärdet av AIC, AICc, eller BIC som är intressant. Utan skillnaden mellan två eller flera modeller (Burnham, 2007).

5.3.2 ARCH och GARCH

“Autoregressive Conditional Heteroskedasticity” (ARCH)-modeller används för att beskriva variansen av feltermen i nuvarande tidsperioden som en funktion av storleken av feltermen i tidigare perioder. Till exempel kan en AR(p) process som i ekvation 5.3.2 antas, men i stället för att vara en vitt brus-process så sägs feltermen variera med tid. Om den sedan representeras som en AR(I)-modell, så sägs feltermen följa en ARCH(I)-process.

Om feltermen i stället beskrivs som

$$e_t = \sqrt{v_t} w_t \quad (5.3.10)$$

Där w_t är i.i.d. med väntevärde 0 och varians 1, och v_t definieras som:

$$v_t = \xi_0 + \xi_1 e_{t-1}^2 + \xi_2 e_{t-2}^2 + \dots + \xi_l e_{t-l}^2 \quad (5.3.11)$$

Följer feltermen ” e ” en “Generalized Autoregressive Conditional Heteroskedastic” (GARCH) process. Med parametrarna k och l (Montgomery, 2016, p. 508).

För att ARCH- och GARCH-modeller ska vara användbara måste datapunkterna ofta först log-transformeras för att normalisera dess varians.

6 Empirisk undersökning och resultat

6.1 Faktoranalys

Innan modelleringen av den slutliga faktormodellen, kommer Principal Components Factoring (PCF) tillämpas för att bestämma hur många faktorer som ska vara med i faktormodellen. Som tidigare nämnt är utgångspunkten för PCF att antalet komponenter/faktorer som ska tas med lika med antalet indikatorer, varför den utförda PCF som följer innehåller fyra komponenter/faktorer.

Med PCF metoden finns det ett antal olika kriterier som kan undersökas för att fastställa antal faktorer. Ett sådant kriterium är att först beräkna egenvärdena för korrelationsmatrisen. Antalet egenvärden som är större än 1 är det antal faktorer som man bör ha med i modellen (Sharma, 1995, p. 117).

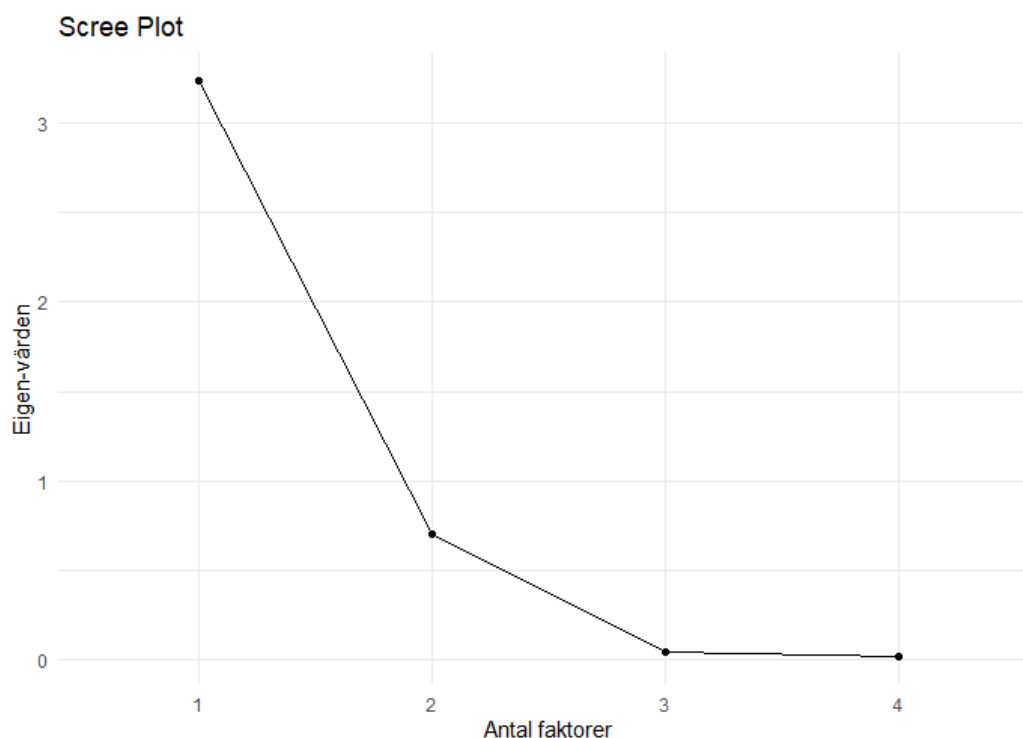
Tabell 6.1.1 Egenvärden från den utförda PCF

Egenvärden	3.236	0.702	0.043	0.019
------------	-------	-------	-------	-------

Enligt utförd PCF korresponderar detta med 1 faktor.

Med egenvärdena kan man också skapa en så kallad scree-plot.

Figur 6.1.1 Scree-plot av egenvärden och antal faktorer



Det värdet på x-axeln då linjen böjer sig är det antal faktorer modellen bör ha med (Sharma, 1995, p. 76). Enligt scree-plotten i Figur 6.1.1 bör modellen innehålla 2 faktorer.

Tabell 6.1.2 Kumulativ varians för de fyra principalkomponenterna.

	PC1	PC2	PC3	PC4
Kumulativ varians	0.809	0.985	0.995	1

Till sist kan man också undersöka den kumulativa förklarade variansen av de olika komponenterna. I detta fall är den kumulativa variansen över 0.8 för en faktor. Utifrån de tre kriterier använda, kan det antas att antingen en enfaktormodell eller en tvåfaktormodell är mest lämplig.

Den påföljande utforskande faktoranalysen har utförts på korrelationsmatrisen av data med PAF-metoden (Se metodavsnittet för en beskrivning av skillnaden mellan PCF och PAF). I utförandet specificeras det att två faktorer ska användas, anledningen till detta beslut förklaras nedan vid examinationen av hur bra faktorlösningarna är.

Tabell 6.1.3 Indikatorernas faktorladdningar

	F1	F2
Område 1	0.900	-0.431
Område 2	0.909	-0.375
Område 3	0.906	0.331
Område 4	0.866	0.494

Rotation av en faktorlösning används ofta när laddningarna för faktorerna är svårtolkade, till exempel om laddningarna för indikatorerna är väldigt lika för de olika faktorerna, i detta fall verkar laddningarna vara tillräckligt distinkta för att någon sådan rotation ej ska vara nödvändig.

I tabell 6.1.3 syns laddningarna för de två faktorerna på de olika indikatorerna. Laddningarna för den första faktorn är höga och positiva. Detta tyder på att denna faktor bidrar positivt till alla indikatorer, ett exempel på en sådan möjlig underliggande faktor kan till exempel vara allmän elförbrukning. Högre elförbrukning, oavsett område, leder till högre elpriser. Den andra faktorn har negativa laddningar för område 1 & 2, medan den har positiva laddningar för område 3 & 4. Denna faktor representerar troligen skillnaden mellan de två geografiska paren av indikatorer. Ett exempel på en möjlig underliggande faktor i detta fall kan till exempel vara efterfrågan av el relativt till produktionskapaciteten. Detta relateras till det tidigare nämnda faktumet att de norra områdena har ett överskott på elproduktion medan de södra har ett underskott. I södra Sverige är efterfrågan på el relativt till elproduktionen mycket högre än i norr, detta leder till relativt högre priser i södra Sverige. I norra Sverige är efterfrågan mindre än produktionskapaciteten vilket leder till lägre priser.

Tabell 6.1.4 Indikatorernas totala kommunalitet (avrundat till 3 decimaler)

Område 1	Område 2	Område 3	Område 4
0.996	0.966	0.930	0.995

Tabell 6.1.5 Indikatorernas unika varianser (avrundat till 3 decimaler)

Område 1	Område 2	Område 3	Område 4
0.004	0.034	0.070	0.005

Kommunaliteterna för de olika indikatorerna är väldigt höga. Detta kan tyda på ett antal olika saker. Som sagt är den totala kommunaliteten den delade variansen mellan en indikator och alla faktorer i modellen. Ett högt värde visar alltså på att faktorlösningen passar data väl. Dock finns det visst skäl för försiktighet då ett alltför högt värde på kommunaliteten kan vara ett tecken på att modellen är "överanpassad" eller "övertädd". Båda dessa koncept är relaterade till att man passar en alltför komplex eller omfattande modell som passar träningsdata väl men som inte presterar väl när det kommer till att generalisera nya data. I detta fall anses det inte vara ett bekymmer eftersom antalet faktorer och antalet indikatorer inte är för stora relativt till storleken av datamaterialet. Försättningsvis eftersom kommunaliteten är hög så kommer den unika variansen av indikatorerna vara låg, vilket betyder att variansen som förklaras av faktorer utanför vår faktorlösning är låg.

För att vidare examinera hur bra faktorlösningen är, kommer faktorlösningens residualer och RMSR att användas. För en god faktorlösning önskas små residualer och ett lågt RMSR-värde (Sharma, 1995, p. 106).

Tabell 6.1.6 Residualmatris för tvåfaktorlösningen.

	Område 1	Område 2	Område 3	Område 4
Område 1	0.004	0.0003	0.001	-0.001
Område 2	0.0003	0.034	-0.001	0.001
Område 3	0.001	-0.001	0.070	0.0001
Område 4	-0.001	0.001	0.0001	0.005

Tabell 6.1.7 RMSR för enfaktor (FA1) och tvåfaktorlösningarna (FA2).

	FA1	FA2
RMSR	0.395	0.034

Residualmatrisen för tvåfaktorlösningen (FA2) visar små residualer vilket tyder på en god faktorlösning. RMSR är lika med 0,034 vilket är ett ytterligare tecken på att två faktorlösningen är lämplig för data och förklarar korrelationen mellan indikatorerna väl.

Tabell 6.1.8 Residualmatris enfaktorlösningen (avrundat till 3 decimaler).

	Område 1	Område 2	Område 3	Område 4
Område 1	0.260	0.213	-0.101	-0.123
Område 2	0.213	0.207	-0.106	-0.110
Område 3	-0.101	-0.106	0.187	0.226
Område 4	-0.123	-0.110	0.226	0.358

Utifrån Tabellerna 6.1.7 och 6.1.9 examineras residualmatrisen och RMSR för enfaktorlösningen (FA1). Dessa visar stora residualer samt ett icke-önskvärt värde på RMSR, därför användes en tvåfaktormodell vid utförandet av faktoranalysen.

6.2 Tidsserieanalys

6.2.1 ARIMA

Som tidigare beskrivit kommer Auto.arima användas för att uppskatta modellerna för varje område. Datamaterialet kommer dessutom begränsas för att se om ARIMA-processer bättre kan beskriva de mönster som finns i data om den dramatiska volatilitet och variation som uppstått de senaste två åren exkluderas.

Till en början kommer analysen att utföras på hela datamaterialet. Med detta ger Auto.arima funktionen följande resultat för de olika områdena:

Tabell 6.2.1 Uppskattade ARIMA modeller för hela tidsspannet

Område 1	ARIMA (4,1,4)
Område 2	ARIMA (4,1,3)
Område 3	ARIMA (5,1,2)
Område 4	ARIMA (5,1,1)

Ett mönster som uppstår är en mindre gruppering av de nordliga och sydliga områdena, men med så höga parametervärden och brist på säsongsterm blir det tydligt att ARIMA inte effektivt kan uppskatta det fullständiga datamaterialet.

Om datamaterialet begränsas för att exkludera information efter juni 2021, då elpriserna började uppvisa större variation och volatilitet, erhålls i stället dessa resultat:

Tabell 6.2.2 Uppskattade ARIMA modeller för begränsade tidsspannet

Område 1	ARIMA (1,1,1) X (0,1,0) [365]
Område 2	ARIMA (1,1,1) X (0,1,1) [365]
Område 3	ARIMA (1,1,2)
Område 4	ARIMA (1,1,1) X (0,1,0) [365]

Med detta går det att se att alla elpriser i alla områden förutom område 3, Stockholmsområdet, kunde uppskattas med samma modell. När data efter juni 2021 fram till april 2023 tas med i analysen syns det i stället att modellerna för de nordliga (1 & 2) och sydliga (3 & 4) områdena mer liknar varandra, samtidigt som att den ökade

volatiliteten minskar effektiviteten av ARIMA-modellering. Därmed minskar de uppskattade modellernas möjlighet att predicera förändringar i elpriser.

Två anledningar till att slutsatsen kan dras att ARIMA-modellerna inte är bästa möjliga modellerna för att beskriva datamaterialet är på grund av att, 1; säsonalitetselementet inte längre syns i de uppskattade modellerna när all data analyseras. 2, på grund av den tidigare nämnda ökningen i volatilitet, vilket kommer skapa icke återkommande anomalier, någonting som underminerar antagandena som måste göras för att ARIMA-modeller ska vara signifikanta.

Viktig information som kan tas från de uppskattade modellerna är att samtliga modeller har d-parametern lika med 1, vilket innebär att de inte är stationära utan att differentieras. Någonting som kommer behöva korrigeras innan GARCH-modellering utförs.

Auto.arima som nämndes tidigare är en funktion i R som automatiskt jämför AIC, AICc, och BIC av modeller med olika parametervärden och returnerar den med bäst anpassning till datamaterial. Denna funktion har använts för att mer effektivt hitta den bäst passade modellen.

6.2.2 GARCH

För att kunna använda Garch-funktionen i R måste den naturliga logaritmen av datamaterialet tas, samt måste differentiering utföras. Efter dessa två steg kan en modell passas.

På grund av att samma Auto.arima funktion inte finns för passning av GARCH-modeller har parameteruppskattningen i stället förts manuellt med hjälp av AIC.

Efter att datamaterialet har anpassats kan de nya GARCH-modellerna jämföras med sina ARIMA-motparter. Syftet denna jämförelse fyller är att skatta modellenanpassning mellan ARIMA och GARCH, inte att göra en direkt jämförelse mellan de tidigare estimerade ARIMA-modellerna med de nya GARCH-modellerna. Resultaten står i tabell 6.2.3.

GARCH-modellerna som har skattats är som följande

Tabell 6.2.3 Passade GARCH-modeller

Område 1	GARCH (6,3)
Område 2	GARCH (5,2)
Område 3	GARCH (5,2)
Område 4	GARCH (2,1)

Tabell 6.2.4 Jämförelse av AIC mellan ARIMA och GARCH modeller passade till differentierad och logaritmerad data

	d.f.	AIC
Område 1 ARIMA	4	1593.6108
Område 1 GARCH	10	721.2594
Område 2 ARIMA	4	1599.6859
Område 2 GARCH	8	727.6541
Område 3 ARIMA	6	2326.1143
Område 3 GARCH	8	1521.8205
Område 4 ARIMA	7	2322.2034
Område 4 GARCH	4	1783.8765

Som illustrerat uppvisar alla GARCH-modeller lägre AIC än sina ARIMA motparter, vilket innebär att GARCH-modeller kommer att vara bättre anpassade till datamaterialet och volatilitet som existerar däri. En reduktion av AIC värden från ca 25% upp till ca 55% är mycket signifikant och menar att ARIMA som möjligt modellerings verktyg tillför väldigt lite av värde till analysen (Burnham, 2007). Frihetsgraderna är inte lika mellan modeller även om de uppskattar samma område. Detta är att förvänta på grund av att frihetsgrader för ARIMA- och GARCH-modeller räknas ut genom summan av parametrarna för modellen, någonting som kommer att skilja sig mellan metoder.

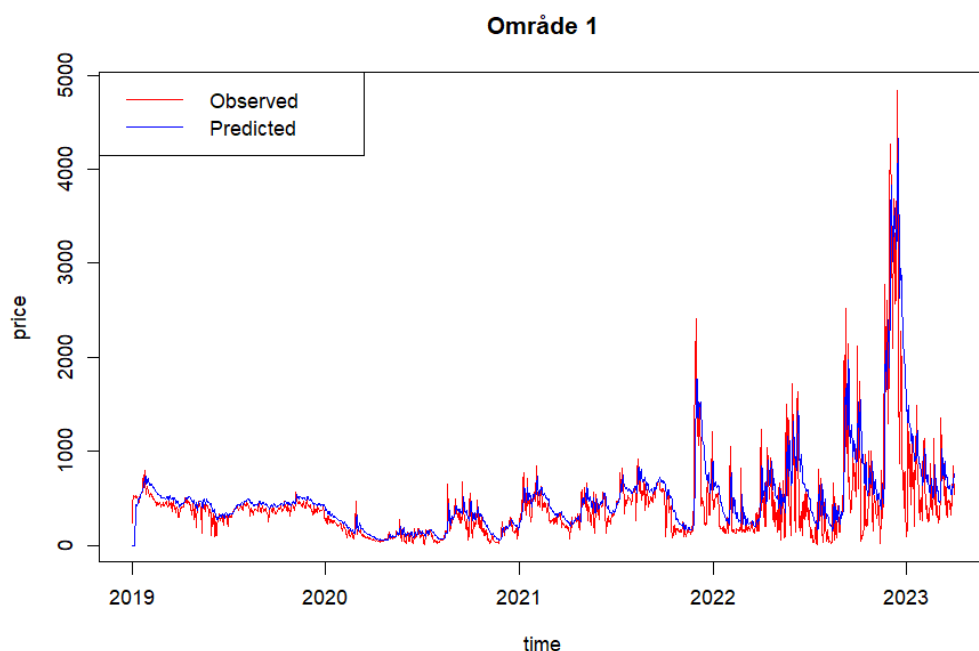
Med modeller uppskattade för alla områden är nästa steg att undersöka hur väl dessa modeller kan användas för att predicera spotpriser på elmarknaden. Som redan etablerats finns det med stor sannolikhet säsongsvariation närvarande i datamaterialet. På grund av detta kan inte GARCH-modellerna heller sägas vara optimala för att uppskatta eller predicera denna sorts tidsseriedata.

6.2.3 Predicering

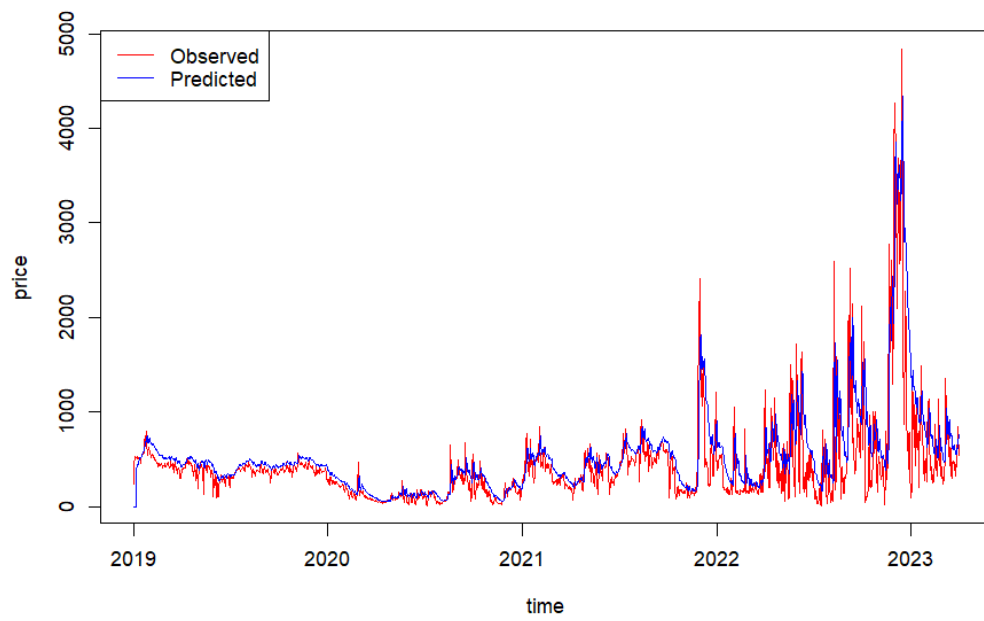
Nästa steg är att använda de uppskattade modellerna för att predicera elpriser för de fyra elområdena. För att kunna utvärdera hur väl modellerna har presterat kommer inte framtida elpriser att prediceras utan i stället datum som redan passerat för att ha en möjlighet att jämföra de predicerade värdena med de observerade.

De modellerna som kommer användas är GARCH-modellerna då de tydligt uppvisade en betydligt bättre anpassningsgrad än ARIMA-modellerna som uppskattades. Hela tidsspannet kommer att användas för prediktionerna.

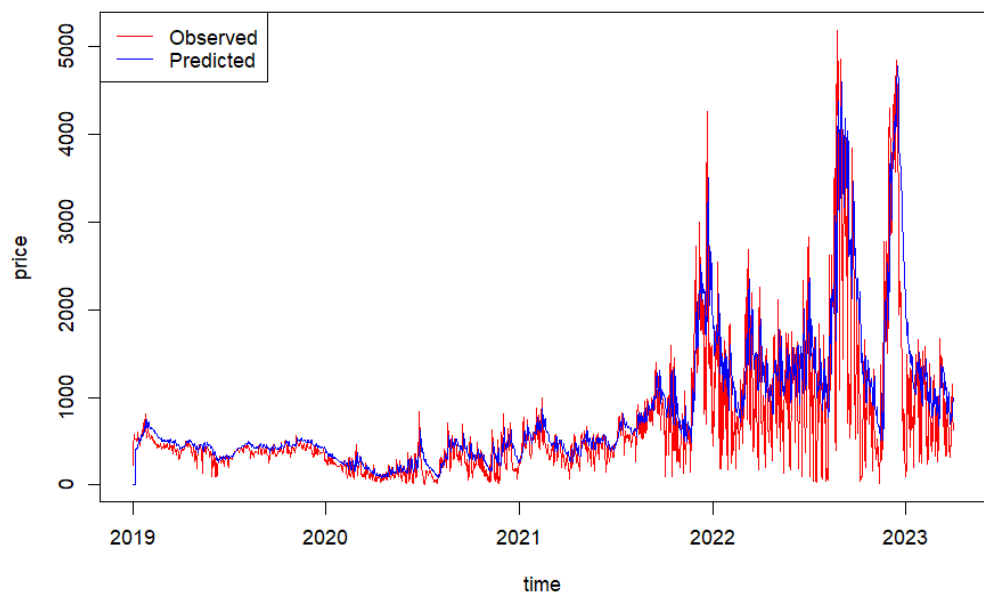
I figur 6.2.1 visas prediktionerna mot de observerade värdena för område 1 till 4 i stigande ordning. Observera hur de predicerade värdena precis följer samma grova mönster som de observerade utan att helt kunna fånga upp topparna och dalarna. Den röda linjen är observerade spotprisivärden och den blåa linjen är de predicerade värdena.

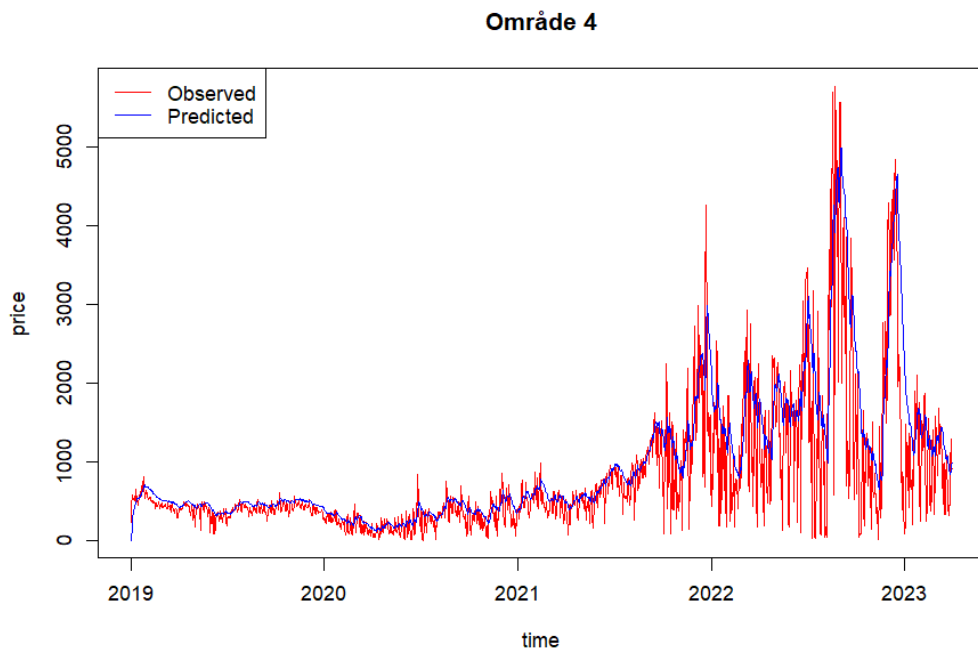


Område 2



Område 3





Figur 6.2.1 Jämförelse av predicerade och observerade elpriser, område 1 – 4 stigande.

En del av variationen har inte kunnat plockas upp av de uppskattade modellerna, men det är tydligt att det underliggande mönstret syns.

I område 3 och 4 som visar tecken på en högre mängd volatilitet mot slutet av observationstiden är det tydligt att modellerna inte kunnat anpassa sin predicering likaväl som till område 1 och 2. Ett förväntat resultat som uppkommer med alla modeller om man tar hänsyn till att modellerna inte ska uppvisa en stor nivå överanpassning.

När Box-Ljung test utförs på residualerna finns signifikant autokorrelation mellan residualerna för samtliga uppskattade modeller. Detta innebär att modellerna med all sannolikhet inte innefattar all information i det ursprungliga datamaterialet. Ett intressant resultat är dock att området med lägst värde av Chi-2 är område 3, Stockholm, vars modell tidigare hade visuellt uppskattats att följa de observerade datapunkterna sämst. Resultaten syns i tabell 6.2.5.

Tabell 6.2.5 Box-Ljung testresultat

	Chi-squared	d.f.	p-value
Område 1	812.2	3	< 2.2e-16
Område 3	699.64	3	< 2.2e-16
Område 3	480.21	3	< 2.2e-16
Område 2	865.38	3	< 2.2e-16

Det finns andra test såsom “Weighted portmanteau test” beskrivet av Li och Mak (1994) som går att utföra, men samtliga paket som erhåller funktionaliteten har fasats ut och finns inte tillgängliga för någon modern version av R.

Den höga nivån av autokorrelation är indikativ på att de prediktioner som skapats inte uppfyller alla krav för att vara en pålitlig representation av spotpriser och de mönster som finns däri.

7 Diskussion och slutsats

De resultaten som har kommit fram visar med stor sannolikhet att modeller ur GARCH-familjen kommer att vara bättre på att predicera elpriser i Sverige än rena ARIMA-modeller, detta är troligen på grund av hur elpriserna har mer sporadiskt fluktuerat under de senaste åren.

Effekten av de fluktuationerna på ARIMA-modellernas möjlighet att representera data syns starkast då datamaterialet begränsas. Specifikt till mellan 2019/01/01-2021/06/01. Innan begränsningen hade de bäst anpassade modellerna ingen säsongperiod, utan i stället mycket höga parametervärden. En gruppering mellan de nord- och sydliga områdena kan också observeras. Efter begränsningen följde alla förutom område 3, Stockholm, samma ARIMA-modell med en säsongperiod av 365.

GARCH-modellerna å andra sidan, som är bättre anpassade till att ta hänsyn till volatilitet, kunde mer effektivt uppskatta elpriser för alla fyra områden oavsett om datamaterialet var begränsat eller inte. När de GARCH-modeller som uppskattats används för att predicera elprisdata med hjälp av de tidigare datapunkterna visas det att modellerna lyckas fånga det generella mönstret av tidsserien för hela tidsspannet.

Ett problem som GARCH-modellerna hade var förmågan att predicera de stora fluktuationerna mellan högt och lågt spotpris som kunde uppstå inom korta tidsspann. En slutsats som bör dras är att precis som i tidigare arbeten så ska inte modellerna som beskrivs här användas eller antas vara mer än ett riktmärke för spotpriset, då de aldrig kommer kunna exakt mäta den reella variationen som priserna upplever.

Under arbetets gång har en mängd utvecklingar av både ARIMA och GARCH som har potential till att bli bättre kandidater stötts på. Exempel på dessa är “Autoregressive Fractionally Integrated Moving Average” (ARFIMA), som beskriver processer där autokorrelation mellan observationer minskar då avståndet mellan dem ökar (Granger, 1980). Eller “Seasonal Fractionally Integrated GARCH” (SFIGARCH), som både tar hänsyn till autokorrelation som minskar mellan observationer och säsonalitet i tidsseriedata (Bordignon, 2005). Den modellen SFIGARCH bygger på och utvecklar, nämligen FIGARCH, som inte tar hänsyn till säsonalitet är också ett användbart alternativ (Tayefi, 2016).

På grund av datamaterialets utförande har FDL, eller andra modeller med fördelad tidsförskjutning, inte kunnat uppskattas eller användas. Detta är på grund av att datamaterialet endast innehåller beroende variabler. Det är sannolikt att modeller med fördelad tidsförskjutning skulle ha varit ännu bättre på att beskriva elpriser än GARCH-modeller, men tyvärr finns ingen vidare möjlighet att utforska det i samband med detta arbete.

Den saknade möjligheten till att inte kunna använda modeller med fördelad tidsförskjutning uppskattas att vara en av de större felkällorna i denna rapport. Antagandet kvarstår att händelser runtom i världen har haft en förskjuten effekt på elspotpriset, och att modellera detta skulle vara intressant att titta på i framtida forskning.

Med hjälp av faktoranalysen har en underliggande struktur som förklarar gemenskapen mellan de olika svenska elområdena kunnat identifieras. Den mest lämpliga strukturen skildras av en tvåfaktormodell, där den första faktorn visade sig vara positivt korrelerad med alla indikatorer. Denna faktor kan antas representera allmän elförbrukning i landet, desto högre elförbrukning desto högre priser oavsett område.

Faktor 2 visar i stället en negativ korrelation med område 1 och 2, men en positiv korrelation med område 3 och 4. Faktor två ska med störst sannolikhet tolkas som en geografisk markör. Detta stärks på grund av att samma mönster visar sig då ARIMA-modellering utförs, och samma gruppering syns mellan områden då samtliga datapunkter räknas med. Denna gruppering är troligen relaterad till elproduktionskapaciteten i de olika områdena. Faktor 2 representerar således produktionskapaciteten relativt till efterfrågan av el. Eftersom Luleå (SE1) och Sundsvall (SE2) har högre elproduktionskapacitet relativt till efterfrågan så leder detta till lägre priser, medan Stockholm (SE3) och Malmö (SE4) har högre efterfrågan än elproduktionskapacitet vilket leder till högre priser.

8 Framtida forskning

På grund av det ständigt förändrande globala landskapet kan endast utbildade gissningar göras kring vad som kommer hända på elmarknaderna närmast. Med det följer en stor möjlighet för forskare och analytiker. En infallsvinkel för framtida forskning kommer troligt baseras på ett mer sociologiskt perspektiv där paralleller dras mellan världshändelser och förändringar i spotpriset för att på så sätt kunna bygga mer robusta modeller med högre reliabilitet.

Problematiken kring modellering av marknadsdata syns tydligt i denna rapport och många andra. Även inom elmarknaden där priser sätts efter reella råmaterial tillgångar är det huvudsakligen spekulation. Att någonsin skapa en modell som kan predicera marknadspriser med målet att erhålla mer vinst är någonting som forskare och matematiker länge har sökt efter. Framsteg har gjorts men en modell som för samtliga fall kommer att kunna predicera vad som händer nästa stund kommer troligen aldrig att uppskattas.

Litteraturförteckning

Adrian Trapletti [aut], Kurt Hornik [aut, cre], Blake LeBaron [ctb] (BDS test code). (2023). Fit GARCH Models to Time Series. rdr.io. <https://rdr.io/cran/tseries/man/garch.html>

Andersson, J., Lillestøl, J. (2010). Multivariate Modelling and Prediction of Hourly One-Day Ahead Prices at Nordpool. In: Bjørndal, E., Bjørndal, M., Pardalos, P., Rönnqvist, M. (eds) Energy, Natural Resources and Environmental Economics. Energy Systems. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-12067-1_9

AS, Nord Pool. (2023). Elspot, <ftp://ftp.nordpoolgroup.com>. Retrieved: 2023/04/01

AS, Nordpool, See outline of our power market history. (n.d.). <https://www.nordpoolgroup.com/en/About-us/History/>

Bordignon, S. (2005). Working Paper Series, N. 18, September 2005 SFIGARCH: a seasonal long memory GARCH model

Burnham, K. P., & Anderson, D. R. (2007). Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach. Springer Science & Business Media.

Damberg, J. (2023, januari) *Därför är elpriserna så höga*. Vattenfall. <https://www.vattenfall.se/fokus/tips-rad/varfor-ar-elpriset-sa-hogt/>

Energimarknadsinspektionen, Så här fungerar elmarknaden. (2021, March 3). <https://ei.se/konsument/el/sa-har-fungerar-elmarknaden>

Fagerholm, C. (2019). Time series analysis and forecasting: Application to the Swedish Power Grid (Dissertation, Linnéuniversitetet). Retrieved from <http://urn.kb.se/resolve?urn=urn:nbn:se:lnu:diva-88615>.

Granger, C. W. J.; Joyeux, R. (1980). "An introduction to long-memory time series models and fractional differencing". *Journal of Time Series Analysis*. 1: 15–30

Gustavsson, A. (2009). Elpriserna på den nordiska elbörsen: Prognosmodellering med hjälp av ARIMA-modeller (Dissertation, Umeå universitet). Retrieved from <http://urn.kb.se/resolve?urn=urn:nbn:se:umu:diva-34820>.

Li, W. K. and Mak, T. K. (1994), On the squared residual autocorrelations in non-linear time series with conditional heteroskedasticity. *Journal of Time Series Analysis* 15(6), 627-636.

Montgomery, D.C., Murat Kulahci and Jennings, C.L. (2016). Introduction to time series analysis and forecasting (2nd edition). Hoboken, New Jersey: Wiley.

Riksdagen, om en elmarknad med konkurrens Proposition 1991/92:133(n.d.). <https://www.riksdagen.se/sv/dokument-lagar/dokument/proposition/om-en-elmarknad-med->

konkurrens_GF03133/html?fbclid=IwAR0IwZWANuTLnNZd3Ju2OTwY2nW1qvnYadbpoSUF800Alh4laSGdvIm_V_Q

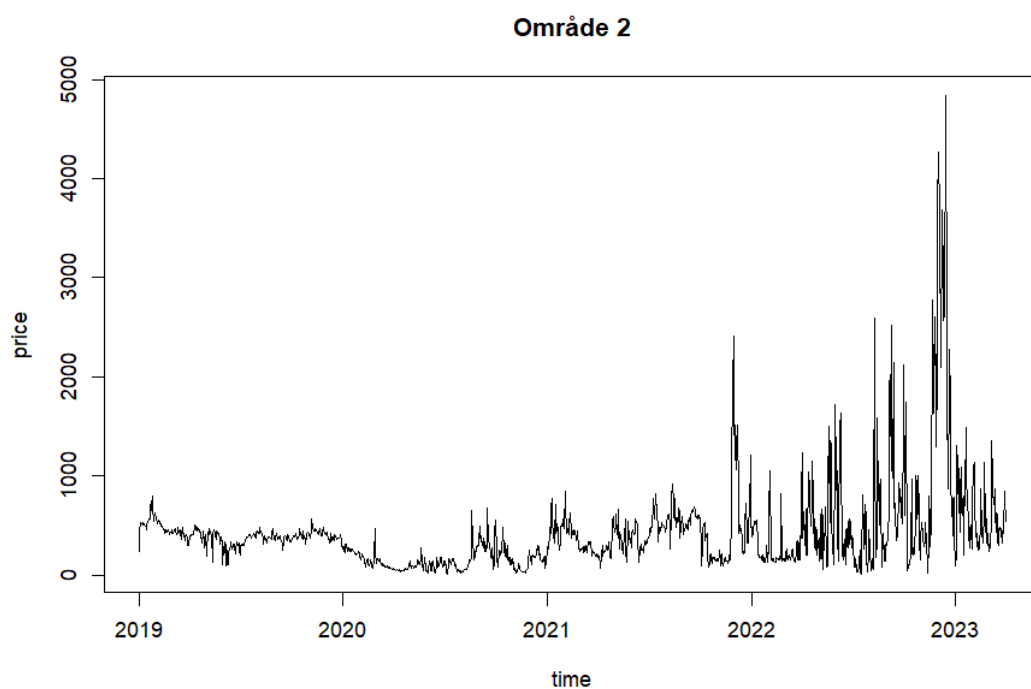
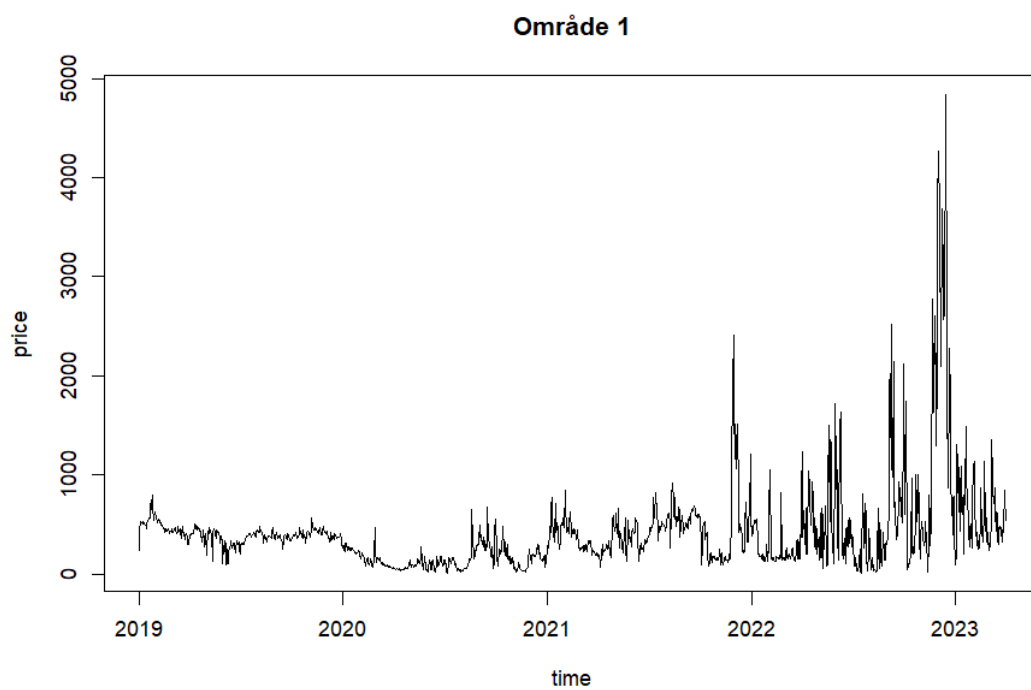
Sharma, S. (1995). *Applied Multivariate Techniques* (1st edition). Wiley, New York.

Svenska Kraftnät. Elområden. (n.d.). <https://www.svk.se/om-kraftsystemet/om-elmarknaden/elomraden/>

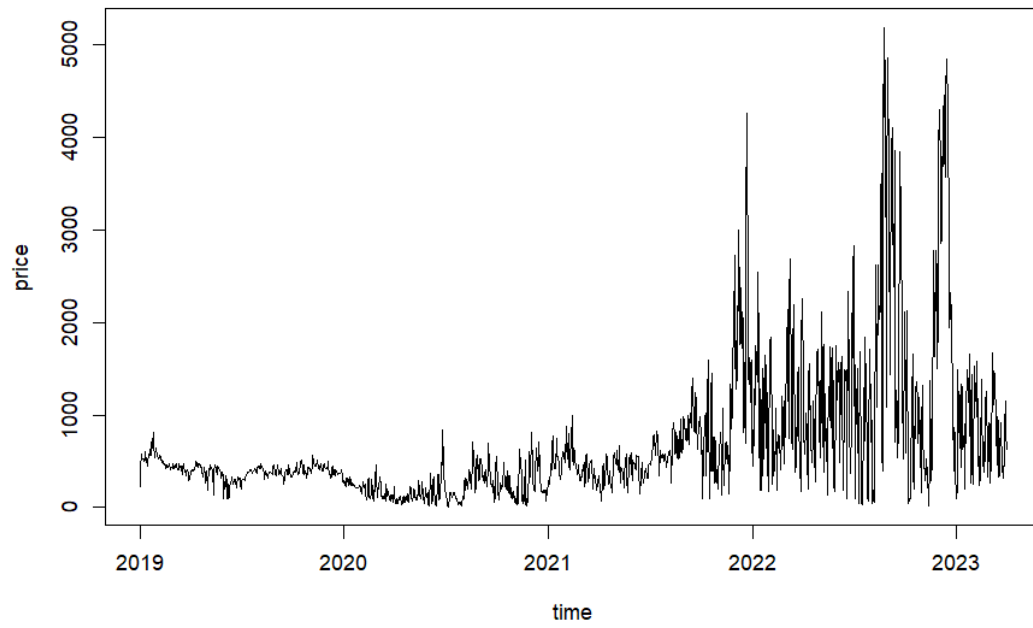
Tayefi, M., & Ramanathan, T. (2016). An Overview of FIGARCH and Related Time Series Models. *Austrian Journal of Statistics*, 41(3), 175–196.

Wooldridge, J. M. (2019). *Introductory Econometrics: A Modern Approach* (5th edition). Cengage Learning.

Bilaga A: Fullständiga tidsserier



Område 3



Område 4

