



LUNDS UNIVERSITET

Statistiska institutionen

# Anpassning av ARIMA-modeller till försäljningsdata

Freddy Andersson & Ellinor Fältman

Uppsats i statistik 15 hp

Nivå 61-90 hp

Maj 2010

Handledare: Björn Holmquist och Dan Jensen

## Sammanfattning

Syftet med denna uppsats är att redovisa resultatet från det uppdrag vi har utfört för GfK, en oberoende part som utför uppdrag från företag som behöver statistisk assistans. Genom tidsserieanalys har vi tagit fram modeller som ska kunna användas för att prognostisera försäljningen för en kund till GfK. Vi erhöll ett 20-tal varugrupper med månadsdata från 2005 till 2008, men har valt att här redovisa sex varugrupper som vi tycker representerar bredden av de metoder vi använt och de problem vi stött på.

Vi har använt oss av Minitab för att bearbeta tidsserierna och bygga modeller. Metoden vi använt oss av bygger på Box-Jenkins metod men med vissa justeringar. Första steget har varit att göra våra tidsserier stationära och även om metoden varierar från varugrupp till varugrupp är där en del gemensam teori vilken har presenterats under kapitel 2 - *metod och teori*. Vi har använt oss av Box-Cox transformation eller log-transformation för att justera för ojämn varians, trend har behandlats med hjälp av avtrendning eller differentiering och där outliers har förekommit har vi använt oss utav en dummyvariabel. Därefter har ARIMA använts för att bygga modeller. Utifrån modellerna har därefter Minitab tagit fram ett års prognos, för varje varugrupp, vilka har bearbetats för att passa den ursprungliga serien.

Vi har funnit att varugruppernas försäljning bäst förklaras med hjälp av säsongmodeller. Av de sex varugrupperna får vi två SAR(1)<sub>12</sub>, tre SMA(1)<sub>12</sub> och en kombinerad MA(1)\*SMA(1)<sub>12</sub>. Anpassningen i modellerna har för det mesta ansetts vara bra då prognoserna ser ut att passa bra till våra ursprungliga serier. Dock har vi stött på några problem under processens gång vilka vi har resonerat kring under kapitel 4 - *diskussion*. Framst skulle vi velat ha haft tillgång till ett större antal observationer, men vi har också diskuterat kring alternativa metoder som skulle kunna gjort vår uppsats bättre. Överlag är vi nöjda med de modeller vi tagit fram utifrån den teori vi tillämpat.

Nyckelord: tidsserie, trend, prognostisering, säsongmodell, transformering, avtrendning, differentiering, försäljning, ARIMA

## **Abstract**

The purpose of this thesis is to display our results received from the assignment GfK addressed to us. GfK is an independent part who aid companies when they need statistical assistance. In order to forecast sales for one of GfK's customers we have build models with time series analysis. We received 26 products with monthly data from 2005 to 2008. In this thesis we have narrowed it down to six of these products, which we find representative in the range of methods we used and the problems we encountered.

We have been using the program package of Minitab when processing the time series and building the models. The method used is for the most part the Box-Jenkins method, but with some adjustments. First step in the process was to make our time series stationary and even though the method alters from product to product, there are some general theories which have been presented in chapter 2. In order to make adjustments for inconsistent variance we have been using Box-Cox transformation or the log-transformation and to process trend we have used detrending and differentiating. Furthermore in cases where outliers are present we have used dummy variables and when building models ARIMA has been our instrument of choice. With these models we finally let Minitab forecast sales for the next coming year.

In our thesis, for the products we have analysed, we find that sales are best explained with a seasonal model. Out of the six products we got two SAR(1)<sub>12</sub>, three SMA(1)<sub>12</sub> and one combined MA(1)\*SMA(1)<sub>12</sub>. The fit of our models has in most cases been good since the forecasts seem to fit well with the original time series. On the contrary we have encountered some problems during the process which has been discussed in chapter 4. First and foremost it would have been favourable if we would have had a larger amount of observations to work with, but we have also discussed alternative methods that could have improved our thesis. All in all we are satisfied with the models built from the applied theory.

Keywords: time series, trend, forecasting, Seasonal models, transformation, detrending, differencing, sales, ARIMA

# Innehållsförteckning

Sammanfattning .....	2
Abstract .....	3
Innehållsförteckning .....	4
1. Inledning .....	5
2. Metod och teori.....	5
2.1 Transformerings.....	6
2.2 Trend .....	7
2.3 Normalitetstest .....	8
2.4 Autokorrelation och partiell autokorrelation .....	9
2.5 Dummyvariabel.....	10
2.6 ARIMA-modeller .....	10
2.7 Ljung-Box .....	12
2.8 Invertibilitet.....	13
2.9 Box-Jenkins urvalsmodell .....	13
2.10 Prognostisering.....	13
3. Analys och resultat.....	15
3.1 Grupp 1 .....	15
3.2 Grupp 2 .....	19
3.3 Grupp 3 .....	22
3.4 Grupp 4 .....	27
3.5 Grupp 5 .....	31
3.6 Grupp 6 .....	34
4. Diskussion .....	37
Referenser .....	39
Bilaga.....	40
Bilaga 1 .....	40
Bilaga 2 .....	43
Bilaga 3 .....	46
Bilaga 4 .....	55
Bilaga 5 .....	60
Bilaga 6 .....	64

# 1. Inledning

Vi har fått i uppdrag av GfK Lund att hjälpa dem genom att försöka få en inblick i hur framtiden för en av deras kunder kan komma att se ut. GfK ska till sin kund ta fram regressionsmodeller för vad som påverkar deras försäljning och vår uppgift är att genom tidsserieanalys bygga passande modeller för att prognostisera försäljningen. Uppgiften inkluderar ett stort antal varugrupper men i uppsatsen kommer vi endast att redovisa sex stycken.

De sex tidsserier vi redovisar representerar i stort sätt de olika scenarion vi påträffat i vårt arbete. Neråt- och uppåtgående trend, varugrupper med deterministisk samt stokastisk trend, tidsserier med tydlig säsongsvarians och tidsserier med mindre tydlig säsongsvarians. Modellerna kommer att byggas utifrån månadsförsäljning (antal sålda enheter) under fyra år, från 2005 till 2008. Vi kommer att inleda denna uppsats med att ta upp relevant teori för att ge en generell bild av hur vi ska gå tillväga. Utifrån detta kommer vi därefter att redovisa metod, resultat samt hålla en kort diskussion för varje varugrupp. Efter att modellerna är gjorda kommer vi även att göra prognoser för försäljningen ett år framåt. Slutligen kommer en mer allmän diskussion och analys att hållas.

Då GfK vill hålla både företaget och vilken bransch det rör sig om hemligt kommer varugrupperna att kallas vid siffror istället för vid namn. Vidare kommer de verkliga försäljningsvärdena inte visas utan dessa kommer att indexeras där januariförsäljningen år ett, 2005, sätts till 100. Det kommer endast att ske en teoretisk analys om hur dessa modeller ska tolkas samt hur förändringar i säsong och trender påverkar modellerna.

## 2. Metod och teori

För att bygga modeller används som utgångspunkt Box-Jenkins strategi. Dock kommer den metod vi presenterar även innehålla andra test och resonemang som vi finner användbara utöver Box-Jenkins. Första steget när en modell ska tas fram är att analysera den plottade tidsserien. På så sätt skaffas en uppfattning över serien och vi kan avgöra huruvida där finns synlig säsongsvarians, tydlig trend eller om residualerna inte har konstant varians över tiden. Det är viktigt att datamaterialet vi sedan bygger en modell utifrån har konstant medelvärde och varians, alltså att tidsserien är stationär. Detta för att vi vid modellkonstrueringen kommer att använda oss av ARIMA-modeller, som antar är att serien är stationär. Generellt sett är inte en tidsserie stationär i sin ursprungliga form utan den kommer att behöva viss bearbetning. Under tiden som bearbetning

görs ska även kontroll av stationaritet utföras efterhand. För att se om tidsserien är stationär undersöker vi autokorrelationsfunktionen och den partiella autokorrelationsfunktionen. Även normalitet kontrolleras löpande med hjälp av en normalfördelningsplott då många test utförs under antagandet att serien är normalfördelad. En annan viktig sak i modellbyggandet och främst vid formandet av de modeller som skall användas till prognostiseringen är att tidsserien förutom att vara stationär också ska vara invertibel.

Om variansen inte är konstant skall vi använda någon form av transformering. Vi kommer att utgå ifrån vad Box-Cox föreslår, men då vi arbetar med ekonomiska tidsserier är log-transformering ett möjligt alternativ. Tidsserier som innehåller trend justeras genom differentiering och/eller avtrendning. I många av våra fall påträffas ett antal outliers. Dessa kan komma att påverka modellbyggandet negativt och kommer vid behov att justeras genom att en dummy-variabel tillsätts. När tidsserien är stationär är det dags att avläsa vilken eller vilka tänkbara modeller som kan komma att förklara serien på bästa sätt och därmed göra bra prognoser. Detta görs genom att analysera autokorrelationsfunktionen och den partiella autokorrelationsfunktionen för att sedan genom ARIMA definiera den bäst anpassade modellen. Utifrån modellen ska vi sedan göra prognoser för ett år framåt i Minitab. Dessa prognoser behöver därefter transformeras tillbaka för att passa den ursprungliga serien.

## 2.1 Transformer

Transformer används för att göra variansen konstant så att tidsserien kan uppnå stationaritet och vi kommer därför att inledningsvis avgöra utifrån ursprungsplotten huruvida variansen är konstant eller ej. I fall vi inte kan avgöra om variansen är konstant kommer vi första differentiera serien för att avgöra om transformering behövs, eller om vi ska använda den ursprungliga otransformerade serien. Vi kommer i huvudsak att använda oss av transformeringen som föreslås av Box-Cox testet. Wei (1994) och Montgomery (1984) beskriver Box-Cox testet som bygger på transformationsparametern  $\lambda$ . Detta värde fås genom att testa olika likelihood-skattningar tills dess att den skattning hittas som minimerar "residual sum of squares". Detta  $\lambda$ -värde motsvarar hur transformeringen ska utföras. Är  $\lambda = 0$  ska serien transformeras genom att ta den naturliga logaritmen av varje värde. Andra transformationsformer uppkommer då Box-Cox testet ger  $\lambda$ -värdena -2, -1, -0.5, 0.5 eller 2. För dessa resultat ska serien upphöjas till det givna  $\lambda$ -värdet. Om Box-Cox testet ger ett  $\lambda$ -värde på 1 behöver inte den undersökta serien transformeras. Enligt en artikel av Nelson & Granger (1979) är det dock bara log-transformeringen som resulterar i normalfördelad data och de menar att Box-Cox inte är det bäst applicerbara vad gäller ekonomiska tidsserier. Då vi vill uppnå normalitet, vilket är en förutsättning för många test som används,

kommer vi ta detta i beaktning, men vi kommer fortfarande att utgå ifrån vad Box-Cox testet rekommenderar.

## 2.2 Trend

Trend är resultatet av stark korrelation mellan tidsseriens observationer. Trend som påträffas kan vara av två olika slag, det finns deterministisk trend och det finns stokastisk trend. Vad gäller deterministisk trend förklarar Cryer (2008) denna som en trend som återfinns då tydliga mönster påträffas i tidsserien. Det kan vara i form av tydligt stigande eller avtagande trend, eller återkommande mönster från säsong till säsong. För att uppnå stationaritet i en tidsserie måste vi justera för dessa mönster. En tidsserie med deterministisk trend görs stationär genom avtrendning. Enligt Enders (2004) kallas trend som det inte ses tydliga mönster i för stokastisk trend. Vid stokastisk trend måste vi differentiera tidsserien för att den ska bli stationär.

### 2.2.1 Avtrendning

Då vi stöter på deterministisk trend kommer vi att använda oss av den avtrendningsmetod som Enders (2004) skriver om. Vid avtrendning används trendlinjen som fås vid trendanalysen. Första steget är att med trendlinjen räkna ut skattade värden för varje tidpunkt som vi har observerade värden för. Därefter subtraheras tidsseriens observerade värden med trendlinjens skattade värden. Serien vi därefter får fram är nu rensad från deterministisk trend och är trendstationär. Om vi vidare finner bevis på att även en stokastisk trend existerar, kommer en första differentiering att utföras. Då en prognostisering görs på en avtrendad serie måste trenden adderas till prognosen i efterhand. Detta görs genom att vi räknar ut skattade värden för de prognostiserade månaderna med trendlinjen och adderar dessa till prognosen. Värt att nämna är att då vi felaktigt behandlar stokastisk trend genom avtrendning istället för att differentiera, resulterar detta i att avtrendningen lägger till en överflödigt trendkomponent, varför det är viktigt att tillämpa rätt metod.

### 2.2.2 Differentiering

Med antagandet att tidsserien inte är stationär på grund av stokastisk trend kan första differensen tas genom  $z_t = y_t - y_{t-1}$  för de observerade värdet  $y_t$  vid tidpunkt  $t$ ,  $t = 2, 3, \dots, n$ . Därefter erhålls den differentierade serien bestående av  $z_1 \dots z_{n-1}$ .

Det är inte alltid en första differens ger ett önskvärt stationärt utseende och då kan en andra differens vara att föredra enligt Bowerman (2005). Denna fås genom  $z_t = (y_t - y_{t-1}) - (y_{t-1} - y_{t-2})$ . Vidare är ett annat alternativ att göra en första säsongsdifferentiering. Denna görs genom  $z_t = y_t - y_{t-L}$  där  $L$  är säsongslängden som säsongsvariationen ska testas på. För månadsdata

används säsongslängden tolv vid ettårig-säsong och säsongslängden sex vid halvårs-säsong. Vidare finns det även möjlighet att göra en 1-12-differens för att uppnå stationaritet, uttrycket som då används är  $z_t = (y_t - y_{t-1}) - (y_{t-L} - y_{t-L-1})$ .

Viktigt att tänka på vid differentiering av en serie är första steget i Box-Jenkins modellurval, vilket behandlas längre fram, som är att en enkel modell är att föredra framför en mer komplicerad. Utifrån denna regel ska differentiering endast göras till dess att den stokastiska trenden är bortrensad. Att då fortsätta differentiera kallas för att överdifferentiera vilket påverkar tidsserien negativt enligt en artikel av Bell (1987). Problemet som uppstår är att överdifferentiering adderar komponenter till ARIMA-modellen samt att antalet observationer i serien minskar för varje differentiering. Utifrån detta kommer vi att ta överdifferentiering i beaktning när vi differentierar. Påträffas en stationär modell med god anpassning vid en första differentiering, kommer inga vidare differentieringar att göras. Om vi felaktigt differentierar en tidsserie med deterministiskt trend kommer ett problem att uppstå i att modellen inte blir invertibel, ett uttryck som förklaras senare.

### **2.3 Normalitetstest**

Med hjälp av Anderson-Darlings test kan vi även undersöka huruvida residualerna är normalfördelade eller ej. Cryer (2008) ger en förklaring av testet som går ut på att jämföra till vilken grad residualernas fördelning överensstämmer med normalfördelningen. Utifrån ett histogram över residualerna fås i vilken frekvens de olika kvantilerna uppgår till. Dessa kvantiler plottas sedan mot de kvantiler som gäller för den teoretiska normalfördelningen, och då materialet är normalfördelat visas de plottade värdena längs en diagonal linje. Den nollhypotes som ställs, att residualerna är normalfördelade, förkastas i fall då p-värdet är litet. Då ett p-värde erhålls som är större än 0.05 anses tidsserien vara normalfördelad.

$H_0$  = Residualerna är normalfördelade

$H_1$  = Residualerna är inte normalfördelade



## 2.4 Autokorrelation och partiell autokorrelation

Autokorrelationsfunktionerna används enligt Cryer (2008) dels för att kontrollera stationaritet och dels för att finna en bra modell till tidsserien. Autokorrelationen visar hur mycket två värden påverkas av varandra. Det finns både positiv och negativ autokorrelation. Positiv autokorrelation definieras av en positiv (negativ) felterm som följs av ytterligare en positiv (negativ) felterm. Negativ autokorrelation definieras av en positiv (negativ) felterm som följs av en negativ (positiv).

Autokorrelationsfunktionen utgår från residualerna som fås vid trendanalysen. För autokorrelationsfunktionen används uttrycket:

$$t_{r_k} = \frac{r_k}{s_{r_k}}$$

Där  $r_k$  är autokorrelationen vid lag  $k$  och  $s_{r_k}$  är autokorrelationens standardfel, tillsammans med gränsvärdet, enligt Minitab:

$$2\sqrt{1 + 2\sum_{k=1}^{i-1} \frac{r_k^2}{\sqrt{n}}}$$

Om  $t_{r_k}$  överskrider gränsvärdet, visar det på en autokorrelation som är signifikant skild från noll vid lag  $k$ . Detta kallas för en spike vid lag  $k$  och om funktionen inte uppvisar fler spikes efter lag  $k$ , har autokorrelationsfunktionen cut-off efter lag  $k$ . En stationär modell har en autokorrelationsfunktion med cut-off på tidiga lags och avtar snabbt mot noll eller i form av en sinusvåg.

Den partiella autokorrelationen bygger på autokorrelation men är rensad för effekterna från de mellanliggande variablerna. Den partiella autokorrelationen mellan  $z_t$  och  $z_{t-k}$  kommer alltså att bero endast på  $z_t$  och  $z_{t-k}$  utan påverkan av  $z_{t-2}$ ,  $z_{t-3}$ , ...,  $z_{t-k+1}$ . Även för den partiella autokorrelationen används uttrycket:

$$t_{r_{kk}} = \frac{r_{kk}}{s_{r_{kk}}}$$

När funktionen ska illustreras men här med gränsvärdet  $\frac{2}{\sqrt{n}}$ , enligt Minitab för signifikansnivån.

Korrelationsfunktionerna används som vägledning för att hitta en bra anpassad ARIMA-modell, detta är något som behandlas senare under stycket *ARIMA-modeller*.

## 2.5 Dummyvariabel

En tidsserie med regelbundna extremvärden, outliers, kan förklaras med en modell innehållande en dummyvariabel som läggs till för den specifika månaden, veckodagen, eller dylikt då seriens outliers inträffar. För handeln är ofta december en månad när försäljningen skjuter i höjden och då en dummyvariabel läggs till kan det förbättra modellen genom att fokus läggs på övriga månader. Vi har för våra tidsserier endast outliers i december månad. När vi nu ska beskriva hur vi går tillväga när serien innehåller sådana kommer vi förklara utifrån att det är decemberoutliers, men metoden går givetvis att använda på vilken outlier som helst.

För en serie utan ökande eller minskande trend, men med outliers i december, är första steget när en dummy-variabel ska adderas att rensa serien för decemberförsäljningarna. Därefter ersätts dessa värden med medelvärdet för alla de övriga observationerna. Med denna nya serie byggs nu en lämplig modell som då förklarar försäljningen från januari till november. En dummyvariabel som motsvarar differensen mellan medelvärdet för decemberförsäljningarna och medelvärdet för övriga serien läggs därefter till. Denna dummyvariabel förklarar nu försäljningen i december.

För serier med en positiv trend tittar vi på varje år för sig. Då ersätts istället varje decembervärde med medelvärdet för respektive års tidigare försäljning mellan januari och november. Dummyvariabeln som därefter ska adderas till modellen är en regressionslinje som beskriver decemberförsäljningens utveckling subtraherat med decemberprognosen som Minitab räknat ut.

Denna metod för att addera en dummyvariabel hänvisas till SCB som använder en liknande metod när deras serier innehåller outliers. Dock är nackdelen att när serien med denna metod jämnas ut kan det resultera i ett uppvisande av autokorrelation där serien inte hade korrelation ursprungligen. Hur mycket autokorrelation som kan uppkomma beror på variationen i extremvärdena. Trots detta väljer vi att använda oss av denna metod då den gör att vi får ut mer information ur serien.

## 2.6 ARIMA-modeller

För att identifiera en passande modell som kan förklara en tidsserie på ett bra sätt utgår vi ifrån den stationära tidsseriens korrelationsfunktioner och tittar på vilken eller vilka lags som är signifikant skilda från noll. Korrelationsfunktionerna visar vilka värden som korrelerar med varandra. En ARIMA-modell av exempelvis ordning två innebär att varje värde baseras på de två tidigare värdena. Det är på detta sätt som ARIMA-modellerna hänger ihop med korrelationsfunktionerna.

En säsongmodell grundar sig på att värdena bygger på varandra med ett visst antal observationer emellan. I detta fall då uppsatsen baseras på månadsdata kommer säsongerna att vara tolv månader.

En ARIMA-modell består av en moving average komponent, kallad MA(q), en autoregressiv komponent, AR(p), samt differentieringsparametern. Modellen skrivs enligt Wei (2006) som ARIMA(p, d, q)(P, D, Q)<sub>s</sub> där p visar vilken ordning den autoregressiva komponenten har och q visar vilken ordning moving average komponenten har. P och Q står för ordningen hos motsvarande säsongmodeller. d och D visar vilken typ av differentiering av serien som ska utföras, för icke-säsongdifferentiering samt för säsongdifferentiering. Slutligen visar s hur lång säsongen, om sådan finns, är uttryckt i antal observationer.

Utifrån ARIMA(p, d, q)(P, D, Q)<sub>s</sub> utvecklas därefter ett uttryck som används vid prognostisering av kommande värden i tidsserien.

$$\Phi_p(B^s)\phi_p(B)(1-B)^d(1-B^s)^D Z_t = \theta_q(B)\Theta_Q(B^s)a_t$$

Detta uttryck är uppbyggt av de olika komponenterna i ARIMA-modellen samt av backshift-operatören B, vilken visar vilka tidigare värden i tidsserien som ska användas.

$$B^j y_t = y_{t-j} \text{ för } j = 0, 1, 2, \dots$$

För differentiering används uttrycken  $(1-B)^d$  och  $(1-B^s)^D$  där d och D vilken typ av differentiering som ska utföras, vanligtvis är dessa lika med ett eller i vissa fall två. Här står s för vilken säsong som används för säsongdifferentieringsparametern. För en modell innehållande en AR-parameter används uttrycket  $\phi_p(B)$  där p motsvarar ordningen för den autoregressiva komponenten. Då modellen innehåller en MA-parameter används uttrycket  $\theta_q(B)$  där q motsvarar ordningen för moving average komponenten. En säsongmodell av autoregressiv typ använder sig av  $\Phi_P(B^s)$  där P visar ordningen på modellen och s står för längden på säsongen. En säsongmodell av moving average typ använder sig av  $\Theta_Q(B^s)$  där Q visar ordningen på modellen.

En vidareutveckling av uttrycken för AR- och MA-parametrarna är följande:

$$Y_t = \Phi_1 Y_{t-1} + \Phi_2 Y_{t-2} + \dots + \Phi_p Y_{t-p} + a_t \quad \text{AR}(p) \text{ modell}$$

$$Y_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q} \quad \text{MA}(q) \text{ modell}$$

Enligt Bowerman (2005) kommer generellt sett en AR(p) modell att vara signifikant på PACF till och med lag p och därefter ha cut off men ha en ACF som går snabbt mot noll. För en MA(q) modell är riktmärket att lags till och med q är signifikant för ACF medan PACF går snabbt mot noll.

Tidsserier från försäljning kan ofta förklaras med en säsongmodell. Som tidigare nämnts så skjuter försäljningen i december i höjden för många branscher men det finns även de som har sommaren som den stora försäljningsperioden. En säsongmodell som SAR(1)<sub>12</sub> eller SMA(1)<sub>12</sub> som grundar sig på månadsdata kommer bygga sina prognosvärden på motsvarande månad för året innan. Vidare kommer en säsongmodell som SAR(1)<sub>6</sub> eller SMA(1)<sub>6</sub> att bygga sina prognoser på försäljningen från sex månader tidigare.

## 2.7 Ljung-Box

För att kunna avgöra om en modell är bra anpassad kommer vi att använda oss av Ljung Box Q-statistics. National institut of standards and technology beskriver testet som testar nollhypotesen vilken är att den anpassade modellens observationer/residualer är slumpmässiga och oberoende, mot att de inte är det. Då nollhypotesen accepteras innebär det att den anpassade serien är oberoende, att lite eller ingen information har gått förlorad i modellbyggandet. Detta syns som ett högt Ljung-Box Q-statistics vilket innebär en bra anpassad modell.

$H_0$  = Residualerna är slumpmässiga och oberoende

$H_1$  = Residualerna är inte slumpmässiga och oberoende

När vi kommenterar detta kommer vi att uttrycka dessa testvärden som Q\*-värden. Alla värden över 0.05 anser vi visa en möjlig anpassning, men med högre Q\*-värden följer bättre anpassning. Det är vidare viktigt att modellen är anpassningsbar för samtliga testade lags, vilket innebär att Q\*-värdena för modellen bör vara höga för samtliga testade lags, lag 12, lag 24 och lag 36.

## 2.8 Invertibilitet

Enligt Enders (2004) är invertibilitet nödvändigt då autokorrelation och partiell autokorrelation fungerar under antagandet att tidsserien kan förklaras av en autoregressiv process. Av denna anledning måste en MA-modell vara invertibel för att prognostisering ska kunna utföras. Med detta menas att moving average processer måste kunna skrivas om till autoregressiva processer. Kriteriet att följa i enlighet med detta är att den slutgiltiga modellens parametrar, ska ha koefficienter vars summa i absolutvärde är mindre än ett. En autoregressiv process är däremot alltid invertibel. På liknande sätt kan en slutlig modell undersökas för stationaritet. En moving average process är alltid stationär men för att en autoregressiv modell ska anses stationär, ska absolutvärdet för parameterens koefficient vara mindre än ett.

## 2.9 Box-Jenkins urvalsmodell

Som vi nämnt tidigare använder vi i stora drag oss av Box-Jenkins strategi i strävan på stationaritet i våra tidsserier. Vidare för att välja bäst anpassningsbara modell kommer vi använda oss utav Box-Jenkins urvalsmetod som Enders (2004) beskriver. Urvalsmetoden sker i tre steg. Först och främst väljs de modeller som är enklast. Fler parametrar innebär mer komplicerade uträkningar. Enligt Box-Jenkins är det då bättre att hålla sig till en modell med något sämre förklaringsgrad om den är av enklare karaktär. Andra steget säger att den modell som uppfyller kraven stationaritet och invertibilitet ska väljas. Detta om modellerna ej kan skiljas åt genom första steget. Slutligen är förklaringsgraden av vikt. Då vi använder Minitab kommer vi att utföra detta steg genom att jämföra Ljung-Box  $Q^*$ -värden och parameterkoefficienternas p-värden som programmet presenterar. Om dessa steg inte räcker till kommer vi utöver Box-Jenkins urvalsmetod även att använda oss av normalitetsplottar och prognoser och dess konfidensintervalls utseenden för att hitta den bäst anpassade modellen. Wei (2006) skriver att breda konfidensintervall är tecken på att processen är icke-stationär, medans smala tyder på stationaritet.

## 2.10 Prognostisering

När en modell som anses vara bra anpassad till tidsserien är identifierad ska en prognostisering göras. Prognostiseringen görs för tolv månader framåt, det vill säga för ett år och görs i ARIMA i Minitab. En prognos kan göras på två sätt, antingen genom forward-prognostisering eller genom backward-prognostisering som Minitab använder sig av. Skillnaden består i att forward-prognostisering prognostiserar framåt med start från det första prognosvärdet medan backward-prognostisering prognostiserar bakåt med start från sista prognosvärdet och därefter

kontrollerar sin prognos genom att gå igenom den framifrån. Enligt Duke University är backward-prognostisering den metod som lyckas använda sig av mest information och är för bra anpassade modeller det bästa sättet att göra ARIMA-prognoser på. Om modellen som ska prognostiseras inte har god anpassning kan detta prognostiseringssätt leda till konvergerande prognoser eller problem med invertibiliteten.

Även ett 95 % -igt konfidensintervall görs efter att prognosen är gjord. Uttrycket som används för prognossteg  $l$  är

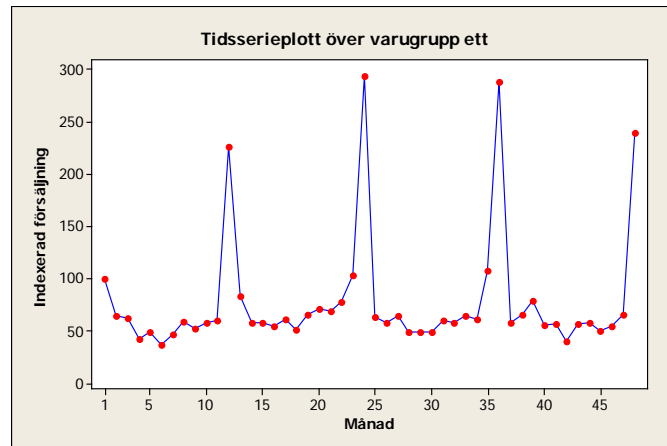
$$\hat{Z}_n(l) \pm 1.6 \left[ 14 \sum_{j=1}^{l-1} \psi_j^2 \right]^{1/2} \sigma_a$$

Vikterna räknas ut med  $\psi_j = \phi^{j-1}(\phi - \theta)$  där  $j \geq 1$  står för differensen som ska utföras i modellen.

### 3. Analys och resultat

Under denna del kommer de sex varugrupperna presenteras var för sig. Vi kommer att gå igenom varje varugrups specifika metod och resultat vid utformning av modell följt av en kort diskussion. Varje varugrupp har en egen bilaga motsvarande sitt nummer där diagram och plottar läggs.

#### 3.1 Grupp 1



Tidsserieplott 1: Varugrupp 1.

Varugrupp 1 är en typisk julklapp. Försäljningen skjuter i höjden i december men har under resterande månader en relativt jämn försäljning. Till denna typ av tidsserie är troligtvis en modell med en dummyvariabel passande. Dummyvariabeln ska förklara decemberförsäljningen och därför tas decemberförsäljningen bort från serien. Dessa värden ersätts med medelvärdet för hela den resterande serien då serien inte har någon uppåt- eller neråtgående trend.

Trendanalys för denna modifierade serie ses i figur 1, bilaga 1. På grund av tydliga skillnader i variansen transformeras denna serie direkt utan extra kontroll med hjälp av första differensen. Box-Cox testet ger ett transformeringsvärde på -0.5, figur 2, och serien transformeras därefter. Transformeringsvärdet räcker dock inte för att göra serien stationär, figur 3, utan även en första differens utförs. Efter transformering och första differentiering verkar serien nu vara stationär, den innehåller ingen trend och variansen är jämn, figur 4. Innan en passande modell bestäms kontrolleras normalitetsplotten samt korrelationsfunktionerna. Normalitetsplotten visar på en serie som ligger strax över gränsen för att anses vara normalfördelad med p-värdet 0.057, figur 5. Både autokorrelationsfunktionen och den partiella autokorrelationsfunktionen går snabbt mot noll, för ACF är lag 1, 7 och 12 signifikanta och för PACF är lag 1 och 7 signifikanta, figur 6 och 7.

Eftersom serien är normalfördelad och korrelationsfunktionerna går snabbt mot noll kan serien anses stationär.

En passande modell ska därefter försöka hittas. Med korrelationsfunktionerna som underlag testas modellerna MA(1), AR(1) och SMA(1)<sub>12</sub>. MA(1), AR(1) och SMA(1)<sub>12</sub> har alla tre signifikanta p-värden för sina koefficienter men väljs bort på grund av att deras Q\*-värden understiger 0,05. Vi har därför testat de kombinerade modellerna AR(1)\*SMA(1)<sub>12</sub>, vilken förkastas då AR(1) komponentens koefficient ger ett ickesignifikant p-värde på 0.063, och MA(1)\*SMA(1)<sub>12</sub> som visar på en bra anpassning. Detta är den slutgiltiga modellen för den här varugruppen. Koefficienternas p-värden är signifikanta med 0.040 samt 0.004 och Q\*-värdena ser bra ut med 0.099, 0.283 samt 0.219. Vidare är koefficienterna 0.3081 och -0.5275 i absolutbelopp mindre än ett, vilket tyder på att modellen förutom stationär också är invertibel, utskrift 1.

Den slutliga modellen är en modell av typen ARIMA(0, 1, 1)(0, 0, 1)<sub>12</sub> och skrivs ut som:

$$(1 - B)Z_t = (1 - \theta B) (1 - \theta B^{12}) a_t$$

Koefficienterna läggs till och modellen skrivs om till ett mer användarvänligt sätt:

$$Z_t = Z_{t-1} + a_t - 0.3 a_{t-1} + 0.85 a_{t-12}$$

Till denna modell ska sedan dummyvariabeln adderas till decembervärdena som plockades bort initialt. Värdet på dummyvariabeln som ska läggas till är differensen mellan decemberförsäljningarnas genomsnitt och genomsnittet för de övriga månadernas försäljning.

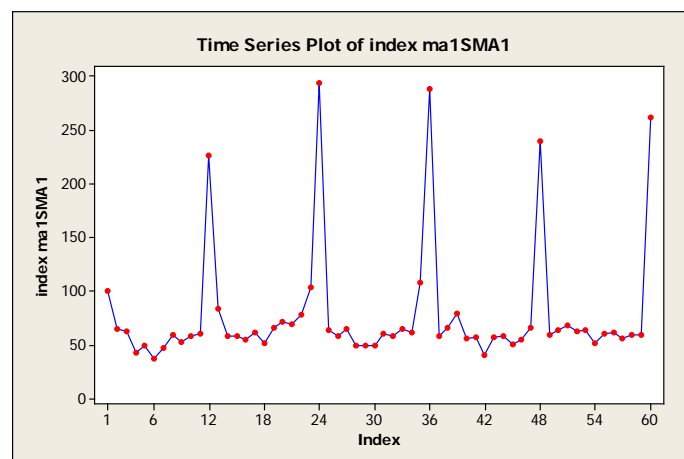
$$\frac{\sum_{d=1}^4 Y_d}{4 \cdot 1} - \frac{\sum_{d=1}^4 Y_j}{4 \cdot 1} + Y_n = 2 - 6 = 2$$

Detta värde ska sedan anpassas efter serien. Då serien är transformerad genom att alla värden är upphöjda med -0.5 så skall även detta värde upphöjas med -0.5. Det ger ett värde på 0.0707. Till modelluttrycket läggs nu dummyvariabeln till där d<sub>12</sub> = 0 för januari till november och d<sub>12</sub> = 1 för december.

$$Z_t = Z_{t-1} + a_t - 0.3 a_{t-1} + 0.58 a_{t-12} + 0.0707 d_{12}$$



Utifrån ovanstående uttryck gör därefter Minitab en prognos genom backward-prognostisering tillsammans med ett 95 %-igt konfidensintervall. Minitab räknar dock fram en prognos för den transformerade serien utan dummyvariabel vilket innebär att prognosen som tas fram måste bearbetas en del. Dummyvariabeln adderas till prognosserien genom att 0.0707 läggs till decemberprognosen. Decemberprognosen har nu inget konfidensintervall men ett sådant skapas genom att förhållandet mellan Minitabs decemberprognos och dess konfidensintervall räknas ut och appliceras på decemberprognosen med dummyvariabel. Nu är hela året prognostiserat men dock är serien i transformerad form i förhållande till den initiala serien. Då ursprungsserien transformerades genom att alla värden upphöjdes till -0.5 ska prognosserien återtransformeras genom att upphöja samtliga prognostiserade värden med -2. På så sätt fås prognosen i samma indexskala som den ursprungliga serien.



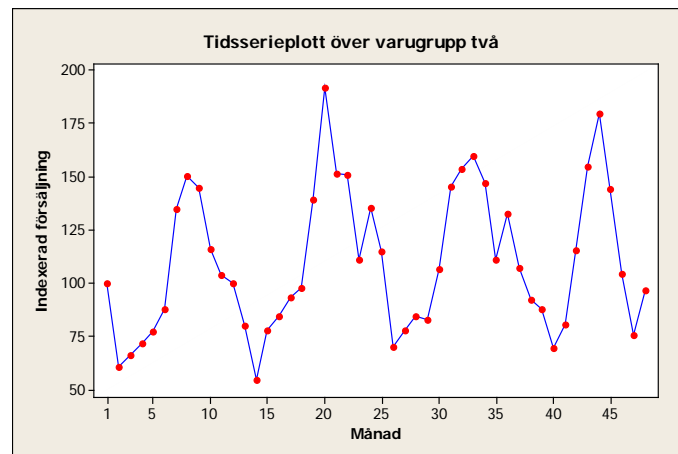
Tidsserieplott 2: Varugrupp 1 samt ett års prognos.

Prognosen ser ut att passa den tidigare försäljningen. Dummyvariabeln är ett genomsnittsvärde av de tidigare decemberförsäljningarna och kommer att bli detsamma för kommande decemberprognoser om vi väljer att göra prognoser för fler år. Prognosen ger små skillnader under året och, förutom för dummyvariabeln, har variansen minskat något. Prognosvärdena ser bra ut med relativt bra konfidensintervall, vilket tyder på att processen är någorlunda stationär, figur 8. Försäljningen kommer med 95 % sannolikhet att ligga mellan dessa värden.

Som noterat skjuter försäljningen för denna varugrupp i höjden under december månad och att behandla dessa outliers med hjälp av dummyvariabel har resulterat i en bra anpassad modell. Att varugruppen kommer att sälja bra i december vet vi redan och det är därför bättre att bygga modellen baserat på informationen vi får från övriga elva månader. Om än något konvergerande, ger modellen en prognos väldigt lik tidigare års försäljning, vars trovärdighet stärks av de bra

konfidensintervallen. Dock uppstår ett problem när det kommer till dummyvariabelns konfidensintervall. Konfidensintervallet blir väldigt brett och framförallt ligger övre gränsen väldigt högt. Vi kommer att stöta på detta senare för varugrupp 4 också, och en diskussion kring detta problem kommer att föras i slutdiskussionen.

### 3.2 Grupp 2



Tidsserieplott 3: Varugrupp 2.

Detta är en varugrupp med stora variationer under året. Försäljningen är högst i augusti och lägst i februari, troligtvis kommer en säsongsmodell vara den bästa. Dock har inte serien någon outlier och en dummy variabel kommer inte att adderas trots stor säsongsvariation.

Huruvida serien behöver transformeras eller ej är något oklart då det är stora skillnader mellan försäljningstopparna och -dalarna men däremot är variansen relativt jämn. Därför tas en första differens för att lättare kunna se om transformering behövs eller ej, figur 1 bilaga 2. Den differentierade serien ser relativt stationär ut, innehåller ingen deterministisk trend och har troligtvis tillräckligt jämn varians för att inte behöva transformeras. Då vi stävar efter att modifiera serien så lite som möjligt testar vi att bygga en modell till serien utan att transformera serien. Därefter kontrolleras normalitetsplotten och korrelationsfunktionerna.

Korrelationsfunktionerna går snabbt mot noll, med lag 12 signifikant för autokorrelationsfunktionen och för den partiella autokorrelationsfunktionen är lag 6 och 10 signifikanta, figur 2 och 3. Normalitetsplotten visar på en normalfördelad serie med p-värdet 0.307, figur 4. Vi anser nu serien stationär och en passande modell ska försöka hittas. Utifrån korrelationsfunktionerna väljs  $SMA(1)_{12}$ ,  $SMA(1)_6$  och  $SAR(1)_6$  som potentiellt fungerande modeller.  $SMA(1)_6$  och  $SAR(1)_6$  ger  $Q^*$ -värden som ligger under 0,05 och väljs därför bort och  $SMA(1)_{12}$  är modellen som väljs. Koefficienten har signifikant p-värde (0,000) och  $Q^*$ -test ger värdena 0.366, 0.323 samt 0.458. Slutligen bör det också nämnas att processen är invertibel med koefficienten -0.7523 som understiger ett i absolutbelopp.

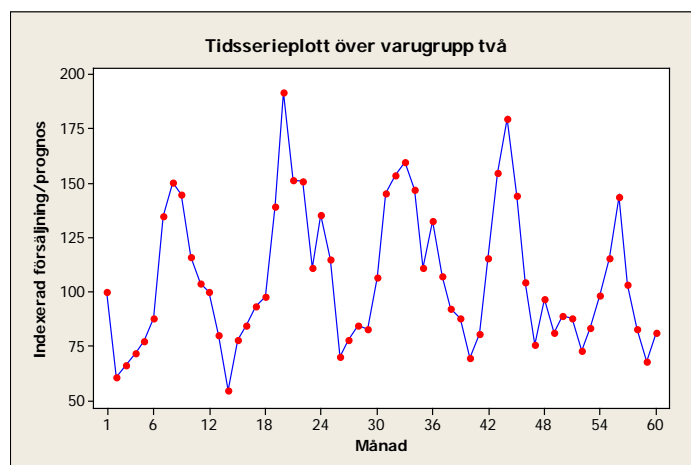
Den slutliga modellen är en modell av typen  $ARIMA(0, 1, 0)(0, 0, 1)_{12}$  och skrivs ut som:

$$(1-B)Z_t = (1-\theta B^{12})a_t$$

Denna skrivs därefter om till en mer användarvänliga formen och koefficienten adderas:

$$Z_t = Z_{t-1} + a_t + 0.7 \cdot a_{t-12}$$

Utifrån ovanstående uttryck gör därefter Minitab en prognos för försäljningen ett år framåt tillsammans med ett 95 %-igt konfidensintervall. Då serien endast är differentierad och ARIMA-modellen innehåller en differentieringsparameter så är de prognostiserade värdena som Minitab levererar i rätt skala direkt.



Tidsserieplott 4: Varugrupp 2 samt ett års prognos.

Prognosen är relativt bra anpassad till modellen men visar på en liten nedgång i försäljningen. Det 95 %-iga konfidensintervallet ger många negativa värden i den undre intervallgränsen, figur 5. Det kan inte en försäljning ha, därför sätts alla de negativa värdena istället till noll, figur 6. Intervallet är fortfarande brett och för exempelvis augusti, som är månaden med störst försäljning, visar intervallet på att försäljningen med 95 % säkerhet kommer att ligga mellan 35 och 255. Detta skulle kunna innebära att vi har att göra med en icke-stationär process, men det är den bästa modell vi hittar utifrån vår metod. Topparna har under de fyra åren legat högre vartannat år. År 5 ska det enligt prognosen vara en lägre försäljningstopp i augusti vilket även stämmer med det tidigare mönstret.

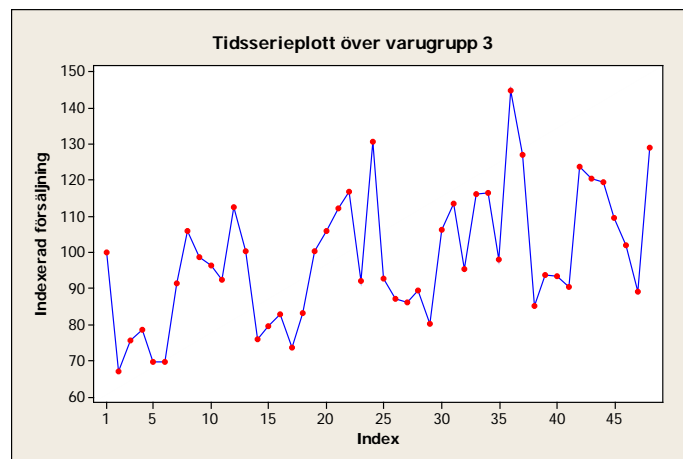
Om vi haft tillgång till en större mängd observationer hade bättre modeller antagligen kunnat tas fram. I ursprungsserien syns bevis på att det förekommer någon form av återkommande ökning vartannat år under sommarperioden. Med försäljningssiffror som sträcker sig längre bak i tiden är

det möjligt att en  $SMA(2)_{12}$  eller en  $SAR(2)_{12}$  vore potentiella modeller som hade gett ännu bättre anpassning.

Dock är problemet med att använda data längre bak i tiden att informationen som ges kanske inte är av relevans. I många branscher bör försäljningen påverkas mer av försäljningen från senast föregående år, än år längre bak i tiden. Detta eftersom marknaden ständigt förändras och produkter kan ersättas av nya.

Vidare kan vi ifrågasätta prognosens trovärdighet då dess konfidensintervall inte är bra. En prognos som säger att försäljningen kommer att ligga mellan gränserna 35 och 255 kan inte anses speciellt givande. Frågan är vad det är som gjort konfidensintervallen dåliga. Det sätt som vår metod för denna varugrupp skiljer sig markant ifrån de andra, är att vi inte transformerat tidsserien då vi ansett att variansen var tillräckligt konstant, kanske skulle den ha transformerats ändå. Att sätta undre gränsen till noll då negativa värden uppkommer, vilket inte är möjligt, påverkar inte konfidensintervallens signifikansnivå och gör ingen skillnad på prognosens trovärdighet.

### 3.3 Grupp 3



Tidsserieplott 5: Varugrupp 3.

Varugrupp tre har en ökande men relativt varierande försäljning mellan månaderna. December säljer alltid bra men dock inte mycket bättre än övriga månader. Försäljningen är något sämre under vårmånaderna och ökar alltid något under sommaren.

Serien innehåller en uppåtgående deterministisk trend och kommer därför att behöva avtrendas. Innan avtrendning görs kontrolleras första differensen då det är något oklart huruvida serien behöver transformeras eller ej, figur 1 i bilaga 3. Då variansen ökar något väljer vi att transformera serien. Vi transformerar Box-Cox test vilket visar  $-0.5$ , figur 2. Trendlinjen vi använder ges i trendanalysen av den ojusterade serien och är  $y_t = 85.4 + 0.455 \cdot t$ . Trendserien räknas ut och därefter transformeras både serien och trendserien. Därefter avtrendas serien och nu visar trendanalysen på att serien innehåller lite neråtgående trend efter avtrendningen vilket gör att vi misstänker att det även finns stokastisk trend i serien, figur 3. För att ta bort trenden första differentieras serien. Därefter ser serien stationär ut, figur 4, och en modell ska försöka hittas. Normalitetsplotten visar på en normalfördelad serie med p-värdet  $0.367$ , figur 5. Korrelationsfunktionerna går snabbt mot noll och signifikanta värden ges på lag 12 för ACF och på lag 2, 5, 7 och 10 för PACF, figur 6 och 7.

Utifrån korrelationsfunktionerna testas  $SMA(1)_{12}$  och  $AR(2)$  men ingen av modellerna ger godkända  $Q^*$ -värden. Eftersom denna modifiering av serien inte gav några bra anpassade modeller måste den justeras ytterligare. Vi tittar på ACF och PACF för den  $-0.5$ -transformerade och detrendade serien, figur 8 och 9. Då tolfte lag för båda är signifikanta så väljer vi att utföra en tolfte differentiering av serien eftersom detta tyder på korrelation mellan säsonger. Efter tolfte

differentieringen analyseras normalitetsplotten vilken visar på ett p-värde på 0.753, figur 10. Därefter undersöks även korrelationsfunktionerna, figur 11 och 12, men då ingen av dessa har lags som är signifikanta finns det inga bra anpassade modeller att hitta.

Vi väljer att byta transformering av serien istället för att göra en andra differentiering då vi anser att den första differentierade serien är stationär och inte innehåller stokastisk eller deterministisk trend och att det hade lett till en överdifferentiering av serien.

Vi använder oss av samma trendlinje som tidigare men innan vi avtrendar så log-transformeras de båda serierna. Efter avtrendningen anser vi att serien fortfarande inte är stationär eftersom det finns viss tendens till stokastisk trend, figur 13, och även en första differentiering måste därför utföras för att serien ska uppnå stationaritet, figur 14. Normalitetsplotten, figur 15, samt korrelationsfunktionerna, figur 16 och 17, undersöks därefter. Serien är normalfördelad med ett p-värde på 0,367. Korrelationsfunktionerna går snabbt mot noll och ACF och PACF visar på signifikanta värden på lag 12 respektive på lag 5, 7 och 10. Utifrån korrelationsfunktionerna testas SMA(1)<sub>12</sub> samt AR(5) men ingen ger bra resultat. SMA(1)<sub>12</sub> ger bra p-värde för koefficienten men dåliga Q\*-värden och AR(5) ger dåliga värden för både koefficienterna och för Q\*-värdena.

En första differentiering av serien gav alltså inga bra resultat. Istället undersöker vi återigen om en tolfte differentiering kan vara ett alternativ. För att se detta tittar vi på ACF och PACF för den odifferentierade serien, figur 18 och 19. För båda korrelationsfunktionerna är lag 12 signifikant och därför görs en tolfte differentiering av serien. Därefter undersöks normalitetsplotten som visar på ett p-värde på 0.768, figur 20, samt ACF och PACF, figur 21 och 22, för den tolfte differentierade serien. Ingen av korrelationsfunktionerna har någon signifikant lag och ingen modell kan ge bra anpassning till serien.

Vi har nu transformerat serien både enligt Box-Cox testet som gav ett transformeringsvärde på -0.5 och genom logaritmering av serien. Därefter har vi provat att både första differentiera och tolfte differentiera de avtrendade serierna. Då ingen av dessa modifieringar har gett någon bra anpassad modell bestämmer vi oss för att frånga vårt ursprungliga antagande om att differentiering behövs. Vi jämför istället de två transformeringarnas avtrendade men odifferentierade trendanalyser för att se vilken som ser mest stationär ut, figur 3 och 13. Ingen av serierna ser ut att vara mer stationär än den andra och vi väljer att titta på normaliteten och korrelationsplottarna för båda.

För den odifferentierade -0.5-transformerade serien är p-värdet för normalfördelningsplotten 0.065. ACF avtar i form av en sinusvåg och har signifikanta lags på 1, 5 och 12, figur 8. PACF går snabbt mot noll och är signifikant på lag 1, 4, 7, 12, figur 9. Utifrån ovanstående resultat testas modellerna MA(1), SMA(1)<sub>12</sub>, AR(1) och SAR(1)<sub>12</sub>. Bäst anpassning ger SAR(1)<sub>12</sub>, utskrift 1, som ger signifikant p-värde för koefficienten (0.000) och Q\*-värden på 0.656, 0.767 samt 0.888.

För den logaritmerade serien gäller liknande resultat. P-värdet för normalitetsplotten är 0.051. Korrelationsfunktionerna ser likadana ut för den logaritmerade serien som för den -0.5-transformerade serien med signifikanta lags för ACF och PACF på 1, 5 och 12 respektive 1, 4, 7 och 12, figur 18 och 19. Även för denna serie testas MA(1), SMA(1)<sub>12</sub>, AR(1) och SAR(1)<sub>12</sub>. Än en gång är det SAR(1)<sub>12</sub> som ger bäst anpassad modell, utskrift 2. Koefficientens p-värde är 0.000 och Q\*-testet ger värdena 0.542, 0.563 samt 0.948.

Nu har vi två modeller som är bra anpassade till serien som i trendanalysen dock inte ser helt stationär ut, men då serien är normalfördelad och korrelationsfunktionerna visar på stationaritet går vi vidare med dessa modeller. Tittar vi på det genomsnittliga Q\*-värdet för respektive modell är det 0.770 för den -0.5-transformerade och 0.684 för den log-transformerade. Box-Cox föreslår -0.5 som transformering men dock är log-transformeringen en enklare transformering som kan vara att föredra. Vi har svårt att välja vilken modell som är den bästa och kommer att titta på båda modellernas prognoser och konfidensintervall i ett försök att finna den bäst anpassade modellen.

I figur 23 och 24 ses prognoserna från båda modeller samt deras konfidensintervall. Prognoserna är mycket lika men serien som är log-transformerad har något smalare konfidensintervall och därför anser vi att detta är den bäst anpassade modellen.

Den slutliga modellen som har bäst anpassning till serien utgår från den avtrendade serien som därefter är log-transformerad men inte differentierad. Modellen är en SAR(1)<sub>12</sub> och uttrycks i andra ord som ARIMA(0, 0, 0)(1, 0, 0)<sub>12</sub>:

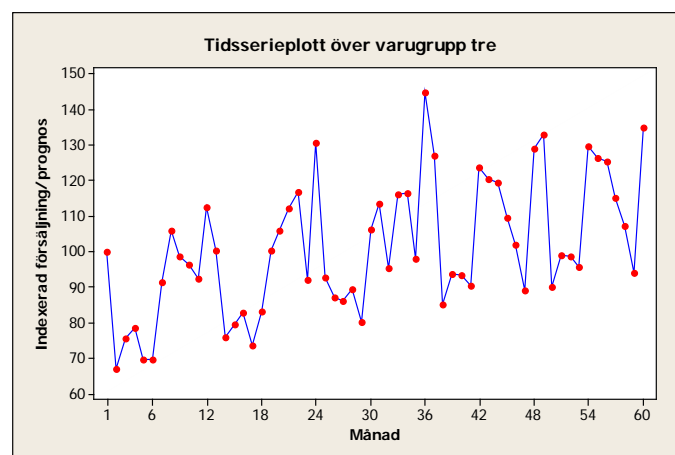
$$(1 - \Phi_p B^s) Z_t = a_t$$

Koefficienten adderas och modellen skrivs om till:

$$Z_t = 0.7 - 3 * Z_{t-1} + a_t$$



Utifrån detta uttryck gör Minitab sedan en prognos för de kommande tolv månaderna tillsammans med ett 95 %-igt konfidensintervall. Minitab räknar dock fram en prognos för den transformerade och avtrendade serien vilket innebär att prognosen som tas fram måste bearbetas en del. Trendkomponenten måste adderas till prognosserien vilket görs med trendserien som användes vid avtrendningen,  $y_t = 85.4 + 0.455*t$ . Prognosvärdena som tas fram är från tidpunkt 49 till 60 och dessa värden läggs till prognosen. Därefter ska den prognosserien med trendkomponenten tranformeras tillbaka så att den passar den ursprungliga serien. Modellbyggandet inleddes med log-transformering och för att transformera tillbaka serien tar vi e upphöjt till prognosvärdena. Detta görs även för intervallgränserna och därefter har vi prognosvärdena i samma indexerade form som den ursprungliga serien.



Tidsserieplott 6: Varugrupp 3 samt ett års prognos.

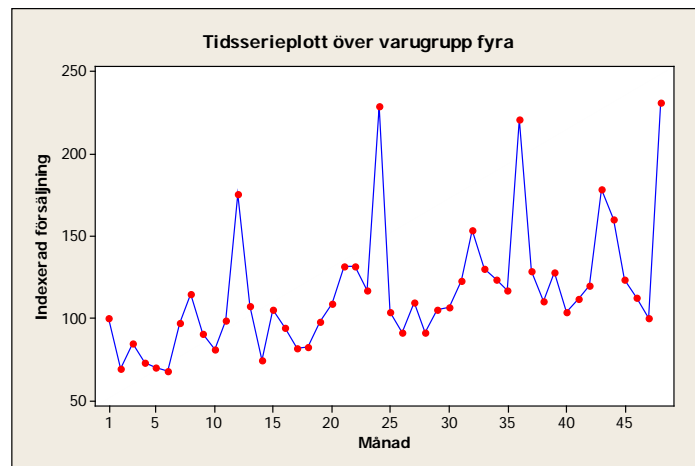
Både prognosen och konfidensintervallet, figur 23, ser ut att vara bra anpassade till serien. Den uppåtgående trenden fortsätter och konfidensintervallet är relativt litet, mellan den övre och den undre gränsen skiljer det sig mellan 30 och 45 enheter. Det smala konfidensintervallet visar på att processen är stationär.

Huruvida Box-Cox bör följas för ekonomiska tidsserier kan diskuteras. Enligt Nelson & Granger är logtransformeringen den transformering som ger mest normalfördelade residualer vad gäller ekonomiska tidsserier. Dock kan vi inte se något samband i vår uppsats. Våra tidsserier som är logaritmerade är inte mer normalfördelade än våra tidsserier som är transformerade med -0.5. Varugrupp 3 är både logaritmerad och -0.5-transformerad. Logaritmeringen gav ett p-värde på 0.051 och -0.5-transformeringen gav p-värdet 0.065 vid test av normaliteten. För denna varugrupp valdes den log-transformerade tidsserien till grund för den slutliga modellen då dess konfidensintervall blev smalare än vad konfidensintervallen för den -0.5-transformerade serien blev. Detta skulle kunna bero på det resonemang Nelson & Granger för om att logtransformeringen resulterar i mest

normalfördelade residualer. Dock är resultatet för normalfördelningsplottarna motsägande där den log-transformerade serien är minst normalfördelad.

Vidare vet vi att tidsserien för de anpassade modellerna fortfarande har någon form av trend. Vi antog först att det fanns stokastisk trend som vi försökte modifiera med differentiering. Dock var vi tvungna att ignorera fenomenet då inga modeller påträffades. Men eftersom vi finner att tidsserien är något så när stationär, residualerna anses normalfördelade och konfidensintervallen är smala, känner vi att vi kan vara nöjda med anpassningen av vår valda modell.

### 3.4 Grupp 4



Tidsserieplott 7: Varugrupp 4

Vid analys av den ursprungliga tidsserien kan vi utläsa outliers i decembermånaderna, men också i slutet av sommaren återfinns försäljningstoppar. Då vi klassificerar decemberobservationerna som outliers rensas serien från dessa toppar. Detta har gjorts genom att byta ut varje års decembervärde mot medelvärdet för respektive års övriga elva månader. Den modifierade seriens trendanalys ses i figur 1 i bilaga 4.

Då tidsserien tycks ha en ökande varians verkar det lämpligt att först transformera serien. Vi tittar även på första differensen av serien som också den visar på ökande varians, figur 2, och därför görs en transformering. Med hjälp av Box Cox får vi ut att den lämpligaste transformeringen är log-transformeringen, figur 3. När vi jämför den logaritmerade serien, figur 4, med den ologaritmerade, figur 1, ser vi nu hur variansen är mer konstant. Dock är serien stigande och har ett återkommande mönster, vilket tyder på påverkan av en deterministisk trend och serien ska avtrendas enligt teorin. Trendlinjen  $y_t = 79.91 + 1.11*t$  fås av trendanalysen för den outliersjusterade serien, figur 1. För att avtrenda logaritmerar vi denna trendserie och subtraherar den från den logaritmerade tidsserien, figur 5. Både autokorrelationen och den partiella autokorrelationen visar på en stationär serie, figur 6 och 7. Normalitetsplotten visar att serien är normalfördelad med ett p-värde på 0.306, figur 8. Trots detta hittar vi ingen modell som passar serien och utifrån detta misstänker vi att det även finns en viss stokastisk trend som vi behöver justera för.

Detta gör vi genom att ta första differensen av serien vilket ger en stationär serie, figur 9. Vid analys av autokorrelationen och den partiella autokorrelationen ser vi att dessa går snabbt mot noll, vilket tyder på stationaritet, figur 10 och 11. Vidare är lag 12 signifikant för ACF och för PACF är lag 3 och 9 signifikanta, men även lag 1 för respektive korrelationsfunktion ligger på gränsen till att vara signifikant, vilket innebär att även denna kan användas som underlag till val av möjlig modell. Normalitetsplotten visar att serien är normalfördelad med p-värdet 0.137, figur 12. Modellerna som testas är SMA(1)<sub>12</sub>, AR(1), MA(1) samt kombinationer av dessa. MA(1)\*SMA(1)<sub>12</sub> ger dåliga Q\*-värden och lika så MA(1). Vad gäller ARMA(1,1) är koefficienten för den autoregressiva parameteren icke-signifikant med p-värde 0.081 samtidigt som Q\*-värdena visar på en dålig anpassning till tidsserien. Även AR(1) anses vara en dåligt anpassad modell till vår tidsserie då dess prognos konvergerar.

Den lämpligaste modellen vi finner är en SMA(1)<sub>12</sub> vilken gav ett signifikant p-värde (0.000) och Q\*-värden på 0.493, 0.330 samt 0.291. Utöver denna modell ger även den kombinerade AR(1)\*SMA(1)<sub>12</sub> en bra anpassning med Q\*-värdena 0.348, 0.293 samt 0.300. Vi väljer SMA(1)<sub>12</sub> eftersom den är av enklare karaktär, men också för att Q\*-värdena är bättre. Vi har också kunnat konstatera att den avtrendade och differentierade tidsserien uppnår, förutom stationaritet, även invertibilitetskravet. Detta då modellens koefficient, som är -0.7631, i absolutbelopp är mindre än 1. Modellen skrivs som ARIMA(0, 1, 0)(0, 0, 1)<sub>12</sub>:

$$(1 - B)Z_t = (1 - \theta B^{-1})a_t$$

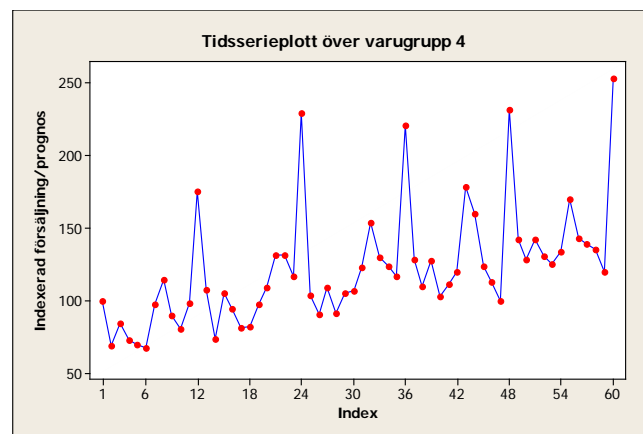
Koefficienten adderas och modellen skrivs i en mer användarvänlig form:

$$Z_t = Z_{t-1} + a_t + 0.7 \quad \text{⑥ } a_{t-1}$$

Utifrån detta uttryck låter vi Minitab göra en tolv månaders prognos för varugruppen. Prognosen som Minitab tar fram presenteras avtrendad och i transformerad form vilket innebär att serien måste bearbetas en del. Därefter ska en dummyvariabel läggas till som ska förklara de bortplockade decembervärdena. Först ska den borttagna trendkomponenten adderas till prognosserien vilket görs med hjälp av trendserien som användes vid avtrendningen,  $y_t = 79.91 + 1.11*t$ . Prognosvärdena som tas fram är från tidpunkt 49 till 60 och det är dessa värden utifrån trendlinjen som läggs till prognosen. Därefter ska prognosen transformeras tillbaka. Transformeringsen som gjordes för serien är log-transformering vilket gör att vi upphöjer e till prognosvärdena för att få prognosen i sin ursprungliga form.

Till prognosen adderas slutligen en dummyvariabel som ska förklara decembervärdena. Då dessa är stigande görs en trendanalys över dessa fyra värden och trendlinjen som fås fram ska användas för att prognostisera kommande decemberförsäljning, figur 13. Trendlinjen är  $Y_t = 17.41 + 5.7x_t$  (där  $x_t$  står för året som ska prognostiseras, försäljningssiffrorna som modellen bygger på täcker år 1 till 4 och år 5 är det vi vill göra prognos över. Dummyvariabeln som ska adderas motsvarar differensen mellan trendlinjen och värdet som Minitab levererar för decemberförsäljningen. För året vi vill prognostisera kommer värdet för dummyvariabeln att vara:

$$Y_d = (17.41 + 5.7 * 5) - 124.5 = 14.5$$



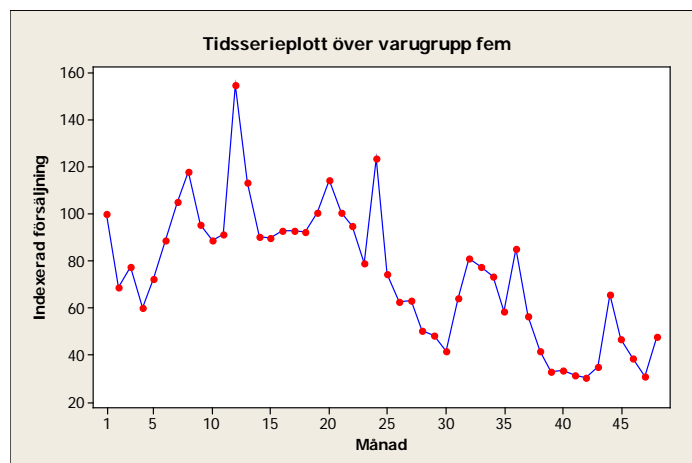
Tidsserieplott 8: Varugrupp 4 samt ett års prognos.

Vidare fås även ett 95 %-igt konfidensintervall ut genom Minitab som med samma uträkningar anpassas till de ursprungliga indexerade värdena, figur 14. Då vi adderar dummyvariabeln till decemberprognosen stämmer inte längre konfidensintervallet som Minitab levererar och vi behöver räkna ut gränserna på egen hand. Först räknar vi ut proportionen mellan prognosen och konfidensintervallen som fås från Minitab. Därefter tillämpar vi denna proportion på dummyvariabeln. Med den här metoden tycker vi att vi löst problemet på lämpligt sätt.

Tidigare i Varugrupp 1 använde vi oss också av en dummyvariabel för att förklara decemberförsäljningen. Då räknades den ut som ett medelvärde av tidigare december värden. Vi tycker att den metod vi använt för att behandla outliers har varit lyckad och antagligen gjort våra modeller bättre än vad de annars skulle ha varit. Decemberförsäljningen står för en stor del av årets försäljning och för att kunna bygga bra modeller som tar hänsyn till hela året är det viktigt att lyfta fram informationen i de övriga elva månaderna.

Även om decemberförsäljningen är väldigt viktig för företag, är det ingen hemlighet att försäljningen ökar runt jul. En avvägning borde då göras huruvida företaget vill prioritera att få en modell som prognostiserar decemberförsäljningen bra, eller övriga elva månaders försäljning bra. Dock tycker vi att de dummyvariabler vi använt till stor del representerar decemberförsäljningen på ett bra sätt., varför vi föredrar att justera serien från decembermånadens outliers.

### 3.5 Grupp 5



Tidsserieplott 9: Varugrupp 5

Den här tidserien har en tydligt neråtgående trend och är ett exempel på en produkt som är på väg ut från marknaden. Vidare ser vi toppar i december månad, men anser att de inte skiljer sig tillräckligt från övriga observationer för att en dummyvariabel ska vara befogad. Variansen är inte jämn nog för att anses konstant och någon form av transformering bör troligen genomföras. För att lättare se huruvida vi ska transformera eller ej tittar vi på första differensen, figur 1 i bilaga 5. Vi ser att den differentierade serien har ojämn varians och därför behöver den transformeras. Vid utförandet av Box-Cox test ser vi att en log-transformering är en lämplig transformering för att göra variansen konstant, figur 2, och då denna transformering är utförd blir variansen något jämnare.

Det syns tydligt i tidsserien att försäljningen är avtagande vilket tyder på deterministisk trend, och detta ska behandlas genom att avtrendera serien. Trendlinjen  $y_t = 108.68 - 1.39628*t$  fås av trendanalysen för den ursprungliga serien, figur 3. Vi logaritmerar trendserien och subtraherar den sedan från den logaritmerade tidsserien, figur 4. Tidsserien blir genom avtrenderingen något mer stationär, men dock inte tillräckligt, och troligtvis innehåller serien även en stokastisk trend, figur 5. Därför väljer vi att gå vidare med första differensen.

Efter en första differentiering kan tidsserien anses stationär, figur 6. Normalitetsplotten ger ett p-värde på 0,130 och serien anses alltså normalfördelad, figur 7. Vid analys av autokorrelationen och den partiella autokorrelationen går båda två snabbt mot noll vilket tyder på stationaritet, figur 8 och 9. Vidare analys ger att signifikanta värden påträffas vid lag 12 för ACF samt lag 6 och 12 för PACF. Med detta som underlag testas anpassningen i de tre potentiella modellerna  $SMA(1)_{12}$ ,  $SAR(1)_{12}$  och  $SAR(1)_6$ .  $SMA(1)_{12}$  och  $SAR(1)_6$  utesluts då de får  $Q^*$ -värden under 0.05. Däremot

ger SAR(1)<sub>12</sub> en bra anpassning med signifikant p-värde (0.000) och Q\* värden på 0.209, 0.295 samt 0.275. Vidare kan vi säga att modellen är stationär då koefficienten som är 0.9860 i absolutbelopp är mindre än 1.

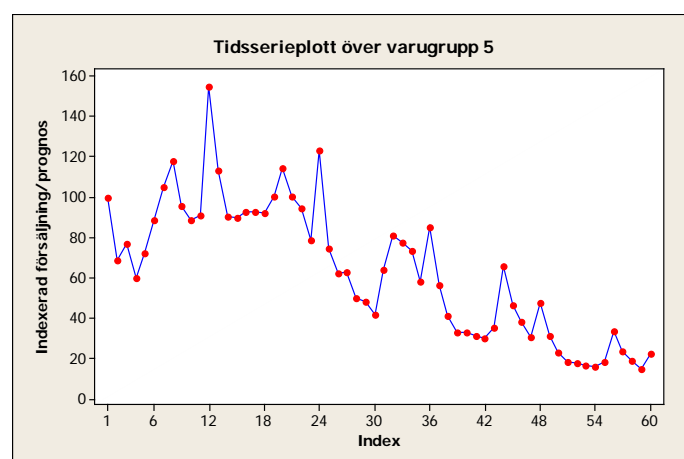
Den slutliga modellen är en modell av typen ARIMA(0, 1, 0)(1, 0, 0)<sub>12</sub> och skrivs som:

$$\Phi_p(B^1) \int 1(-B)Z_t = a_t$$

Denna skrivs sedan om till en mer användarvänlig form och koefficienten adderas:

$$Z_t = Z_{t-1} + 0.986 Z_{t-1} + a_t$$

Utifrån ovanstående uttryck gör därefter Minitab en prognos för tolv månader tillsammans med ett 95 %-igt konfidensintervall. Minitab räknar dock fram en prognos för den transformerade och avtrendade serien vilket gör att prognosen som fås fram måste bearbetas en del. Första steget är att addera trendkomponenten som togs bort i inledningen. Med samma trendlinje,  $y_t = 108.68 - 1.39628 * x_t$ , räknas trenden för varje prognosvärde,  $t = 49, 50, \dots, 60$ , ut och adderas till prognosen. Då ursprungsserien transformerades genom att alla värden logaritmerades ska prognosserien återtransformeras genom att upphöja e med varje prognosvärde. På detta sätt fås prognosen i samma indexerade skala som den ursprungliga serien. Samma bearbetning görs för intervallgränserna vilka är väldigt smala, vilket tyder på en stationär process, och stärker prognosens trovärdighet, figur 10. Nedan visas den ursprungliga indexerade tidsserien tillsammans med de indexerade prognosvärdena.



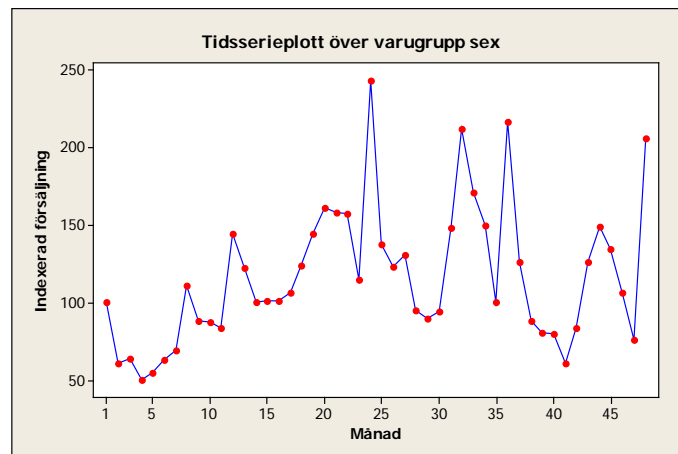
Tidsserieplott 10: Varugrupp 5 samt ett års prognos.



Som syns i plotten ovan, prognostiserar Minitab en fortsatt neråtgående försäljning för varugrupp fem. Antagligen kommer en produkter med denna sortens prognos att försvinna från marknaden, och kanske ersättas av andra produkter. Trots att vi tycker att den valda modellen förser oss med en bra prognos kan vi ifrågasätta nyttan av att modellera en varugrupp som tillsynes är på väg utifrån marknaden. Det skulle vara för att tillämpa modellen på de produkter som ersatt produkt 5. När dessa produkter kommer till slutskedet i sin livscykel och är på väg att ersättas av ännu bättre produkter befinner de sig i ett scenario liknande det för produkt 5.

Vad som talar emot detta begränsade användningsområde är att olika sorters produkter påverkas olika mycket då de blir ersatta på marknaden. I vissa fall, blir den gamla troligtvis inte utbytt till den nya förrän den gamla har gått sönder. I andra fall kommer konsumenterna att efterfråga den nya produkten i större utsträckning, till exempel på grund av nya funktioner, modernare utförande och så vidare, vilket leder till att den gamla produkten byts ut i ett tidigare skede.

### 3.6 Grupp 6



Tidsserieplott 11: Varugrupp 6

Den här varugruppen ter sig lite annorlunda än många andra. Det är svårare att analysera plotten och utifrån det tänka sig hur prognosen kan komma att se ut. Materialet innehåller ett par observationer som kan anses vara starka kandidater till att behandlas som outliers. Dock låter vi bli att retuschera serien eftersom de inte följer ett specifikt mönster. Försäljningen har de första två åren varit ökande men utan stora variationer mellan månaderna. Andra halvan av serien är däremot svagt avtagande med stora variationer under året. Vidare ser vi tydligt att variansen inte är konstant, och någon form av transformering kommer att vara nödvändig.

Genom att utföra Box-Cox analys får vi fram ett transformeringsvärde på  $-0.5$ , figur 1 i bilaga 6. Vi följer Box-Cox och efter transformeringen är variansen jämnare, dock finns det även trend i serien vilken vi misstänker vara stokastisk, figur 2. För att bli av med trenden tar vi första differensen, vilket resulterar i att tidsserien nu kan anses stationär, figur 3. Genom att studera normalitetsplotten ser vi att residualerna är normalfördelade med p-värdet 0.102, figur 4. Vid analys av ACF och PACF ser vi att lag 12 är signifikant för båda korrelationsfunktionerna och med detta som underlag testas en  $SMA(1)_{12}$  och en  $SAR(1)_{12}$  modell. Vidare ser vi också att ACF och PACF går snabbt mot noll vilket stärker vårt antagande att tidsserien gjorts stationär, figur 5 och 6.

Modellen vi bestämmer oss för är  $SMA(1)_{12}$ . Vid test av  $SAR(1)_{12}$  finner vi att koefficientens p-värde uppvisar signifikans, men modellen förkastas då  $Q^*$ -värdena är under 0.05.  $SMA(1)_{12}$  däremot visar signifikant parameter med p-värdet 0.001. Dessutom är  $Q^*$ -värdena signifikanta, 0.875, 0.165 samt 0.338, vilket tyder på att informationen i tidsserien är väl utnyttjad och modellen kan anses ha en god anpassning till den ursprungliga serien, utskrift 1.

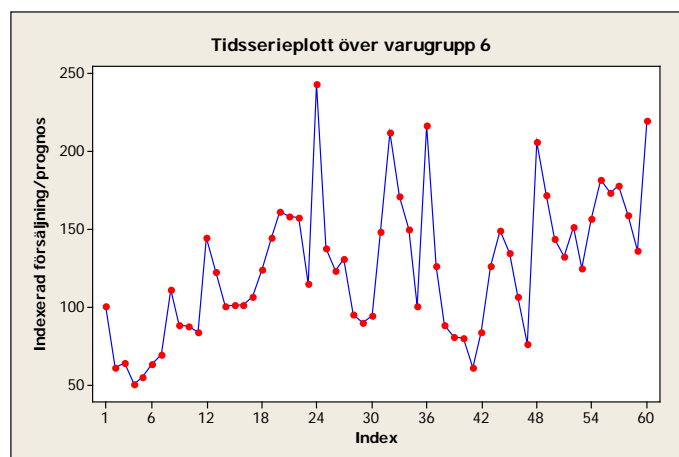
Den slutliga modellen är en modell av typen  $ARIMA(0, 1, 0)(0, 0, 1)_{12}$  och skrivs ut som:

$$(1 - B)Z_t = (1 - \theta B^{12})a_t$$

Därefter adderas koefficienten och modellen skrivs om till:

$$Z_t = Z_{t-1} + a_t + 0.5 \sum_{i=1}^{12} a_{t-i}$$

Utifrån ovanstående uttryck gör Minitab en prognos tillsammans med ett 95 %-igt konfidensintervall för det kommande året genom backward-prognostisering. Minitab räknar dock fram en prognos för den transformerade serien vilket innebär att prognosen som tas fram måste bearbetas. Då ursprungsserien transformerades genom att alla värden upphöjdes till -0.5 ska prognosserien återtransformeras genom att upphöja samtliga prognostiserade värden med -2. På så sätt fås prognosen i samma indexskala som den ursprungliga serien.



Tidsserieplott 12: Varugrupp 6 samt ett års prognos.

Prognosen för varugrupp 6 ser något hög ut. Konfidensintervallets övre gräns ligger väldigt högt och för illustrationssyfte har vi bortsett från den gränsen, figur 7, och tittar istället på prognosen med endast den undre gränsen, figur 8. Undre intervallgränsen säger i sig inte heller väldigt mycket om prognosens trovärdighet, men nuvarande prognos är den bästa vi kan få fram med utgång ifrån den teori vi tillämpat. Likt tidigare för varugrupp 2 har vi här antagligen att göra med en ickestationär process då konfidensintervallen är väldigt dåliga.

Slutligen kan tilläggas att då varugruppen transformerats med -0.5 bestämde vi oss för att även testa för en modell med log-transformation, då detta är det mer vanliga för ekonomiska tidsserier. Dock påträffas ingen modell som ger bättre anpassning än den vi nu valt. Detta är en av de få varugrupper där vi inte hittat någon riktigt bra modell för att beskriva serien, vilket resulterat i en sämre prognos.

Det kan för denna varugrupp diskuteras huruvida vi använt oss av rätt metod vid modellbygget. Ursprungliga tidsserien har ett mer kvadratisk utseende och genom vidare studier hade vi kanske kunnat finna en alternativ metod för den här varugruppen. Det är väldigt oklart hur försäljningen kommer att bete sig i framtiden när vi analyserar tidsserien. Vi vet inte om varugruppen nyligen blivit introducerad på marknaden med en stor försäljning som följd som nu håller på att jämnas ut eller om den är på väg ut från marknaden. Med fler efterkommande försäljningsvärden skulle det antagligen gå att bygga en bättre modell. Den alternativa metoden vi kunde använda för att försöka bygga en bättre modell med den information vi har vore att, liksom avtrendningen med linjär trend, använda en kvadratisk trendlinje för att avtrendera denna. Detta skulle då resultera i en prognos som tyder på en produkt på utgång, men detta kan vi inte säkert veta med den ringa informationen vi har för denna varugrupp.

## 4. Diskussion

Vi anser att vi överlag har lyckats hitta modeller med bra anpassning och som har gett bra prognoser. Dock har vi stött på en del hinder, somliga vi inte har kunnat påverka men några som vi hade kunnat göra något åt. I detta avsnitt kommer vi att diskutera uppsatsen, reflektera över alternativa tillvägagångssätt samt ge förslag på möjliga förbättringar.

En faktor vi inte har kunnat påverka är det ekonomiska läget som har varit ostabilt den senaste tiden efter krisen i slutet på 2008. Vårt data motsvarar försäljningen från januari 2005 till december 2008 och under denna period syns troligtvis inte den nuvarande lågkonjunkturen speciellt mycket. Däremot kommer våra prognoser att grunda sig på en period i högkonjunktur när vi istället prognostiserar försäljning för ett år i lågkonjunktur. Detta innebär att de prognoser vi har gjort, utifrån de modeller vi har tagit fram, med stor sannolikhet kommer att vara högre än den faktiska utgången för försäljningen.

En annan faktor vi inte har kunnat påverka är den ringa mängden observationer som har varit till vår nackdel, vilket avspeglar sig i bland annat breda konfidensintervall för vissa varugrupper. Hade vi exempelvis haft tillgång till månadsdata för en längre period hade våra modeller kunnat ge mycket mer precisa prognoser. Vi hade då kunnat ta i beaktning hur tidigare lågkonjunkturer har påverkat försäljningen och byggt modeller med hänsyn till detta. Vidare är det i regel vanligt att vid modellbygge lämna några observationer till att ha något att jämföra den kommande prognosen med, och återigen uppstår problemet med att vi har för få observationer. För oss skulle det vara att föredra att lämna de tolv sista observationerna för jämförelse då vi arbetar med månadsdata och vill prognostisera ett år, men eftersom vi endast har 48 observationer skulle det innebära att modellbyggandet endast skulle få 36 observationer som underlag. I vår mening är detta för lite information att gå på. En annan väg till att öka observationerna, förutom att använda sig av fler år, vore att använda sig utav vecko- eller dagsdata. Det hade kunnat ge bättre modeller då variansen inom månaderna hade blivit tydligare om en produkt exempelvis säljer bättre i början av månaden än i slutet. Dock hade troligtvis dessa kortare analysperioder gett modeller av mer komplicerad karaktär än de nuvarande  $SMA(1)_{12}$  och  $SAR(1)_{12}$  som genomgående används. Vilken metod som är att föredra är upp till användarna av modellen beroende på hur detaljerad information de vill ha och hur komplexa modeller de vill arbeta med.

Vad gäller de prognoser vi har gjort är de oftast väldigt lika det fjärde och sista året i den ursprungliga tidsserien. Detta faller sig naturligt då vi arbetar med SAR(1)<sub>12</sub> och SMA(1)<sub>12</sub> modeller vilka innebär att det prognostiserade värdet endast beror på försäljningen under motsvarande månad föregående år. En produkt som visar på en stigande försäljning kommer att resultera i en prognos som ligger högre än föregående säsong, och på samma sätt kommer en varugrupp med fallande försäljningssiffror fortsätta visa på nergång i försäljningen. Dock kan en ökande försäljning inte öka för all framtid. Att använda de fyra år av datamaterial vi tilldelats och utifrån detta prognostisera varugruppernas försäljning för en längre tid framåt är omöjligt. Vi utför endast ett års prognos och det resulterar i många fall i en någorlunda bra anpassad prognos. Dock måste modellerna uppdateras i takt med att månaderna går för att bästa möjliga prognos ska kunna göras.

I varugrupp två och fyra stöter vi på problem med att bestämma konfidensintervall för decemberprognosen då den skattas med hjälp av en dummyvariabel. För dess konfidensintervall har vi gjort en intervallskattning proportionellt mot det konfidensintervall som Minitab levererar för det prognostiserade decembervärdet innan dummyvariabeln är tillagd. Ett alternativt tillvägagångssätt för att addera en dummyvariabel och på köpet få ett konfidensintervall hade varit genom att bilda en ny tidsserie endast bestående av de fyra decembervärdena. Utifrån denna tidsserie hade vi kunnat ta fram en modell som prognostiserar försäljningsvärdet för december med tillhörande konfidensintervall. På så sätt skulle dummyvariabelns skattade konfidensintervall kunna ha varit smalare eftersom decemberförsäljningen då hade skattats med en enstegs prognos istället för med en tolvstegs prognos.

Ett annat problem vi hade kunnat behandla på ett annorlunda och måhända bättre sätt är den deterministiska trenden. Vi valde att avgöra utifrån tidsserieplottarna huruvida serien innehöll denna typ av trend eller inte då Minitab inte presenterar någon funktion som testar för trend. Detta test hade däremot kunnat utföras i något annat program, exempelvis Evievs. Vi hade också kunnat prova att justera samtliga tidsserier för deterministisk trend och hade modellerna blivit bättre hade tidsserien troligtvis innehållit denna sortens trend.

Vi har funnit att varugruppernas försäljning bäst förklaras med hjälp av säsongmodeller. Detta faller ganska naturligt då försäljningskampanjer och reor infaller samma tid varje år. Överlag är vi nöjda med de resultat vi fått och de modeller vi tagit fram. Även om den ekonomiska krisen kan utgöra ett problem för våra modeller, som kanske bör tillämpas under perioder då marknaden är mer stabil, tycker vi ändå att uppsatsen kan ligga som underlag till att prognostisera försäljning.

## Referenser

Bell, W., *A note on overdifferencing and the equivalence of seasonal time series models with monthly means and models with (0,1,1)<sup>12</sup> seasonal parts when  $\theta=1$* , Journal of Business and Economic Statistics 5, 1987

Bowerman B., O'Connell R., Koehler A., *Forecasting, time series and regression*, 2005, 4:de uppl., Thomson

Cryer, J. D., *Time series analysis*, 2008, 2:a uppl., Springer

Duke University, *Estimation of ARIMA models*, Uppdaterad 2005-05-16,  
<http://www.duke.edu/~rnau/arimest.htm> (10-04-27)

Enders, W., *Applied Econometric Time Series*, 2004, 2:a uppl., Wiley

Minitab version 15, help function, under: Contents – Methods and formulas – Time series

Montgomery, D., *Design and analysis of experiments*, 2:a uppl., 1984, Wiley

Nelson & Granger *Experience with using the Box-Cox transformation when forecasting economic time series*, Journal of econometric 10, 1979

NIST (National Institut of Standards and technology), *Ljung-Box test*, Alan Heckert, Uppdaterad 2003-04-04, <http://www.itl.nist.gov/div898/software/dataplot/refman1/auxillar/ljungbox.htm> (10-05-05)

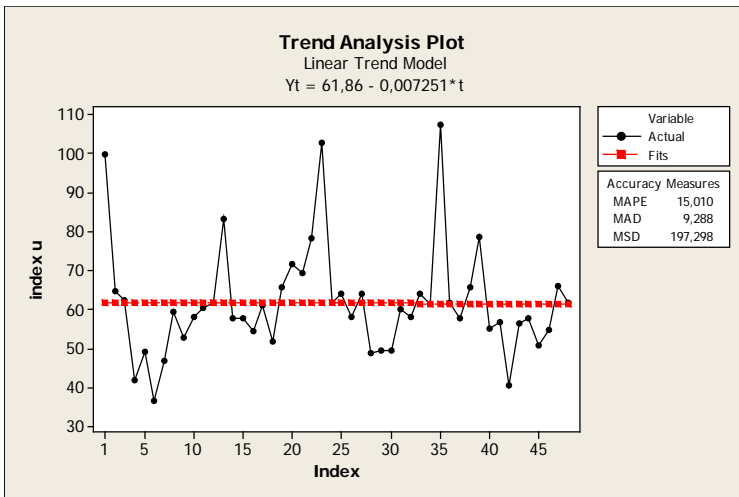
SCB, *Tidsserieanalys av svenska BNP-revideringar 1980-1999*, Nilsson, J., Rosander, O., 2004-06,  
[http://www.scb.se/Pages/PublishingCalendarViewInfo\\_259923.aspx?PublObjId=1967](http://www.scb.se/Pages/PublishingCalendarViewInfo_259923.aspx?PublObjId=1967) (10-04-26)

Wei, W., *Time series analysis – Univariate and multivariate methods*, 1994, 1:a uppl., Pearson Education, Eastbourne

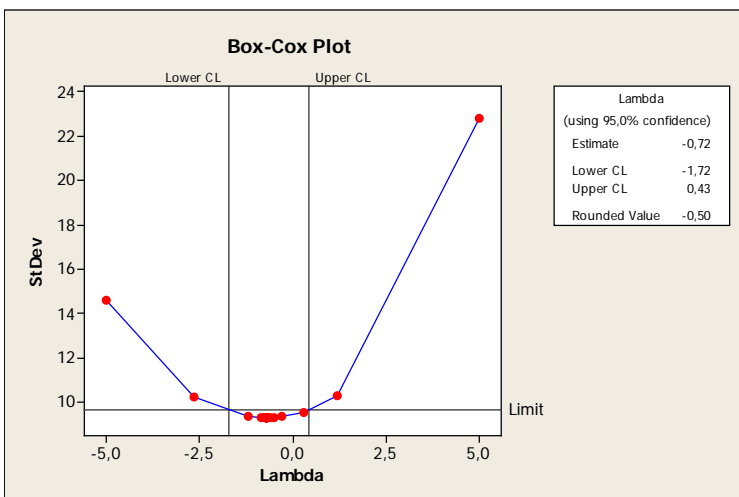
Wei, William W.S., *Time series analysis – Univariate and multivariate methods*, 2006, 2:a uppl., Pearson Education

# Bilaga

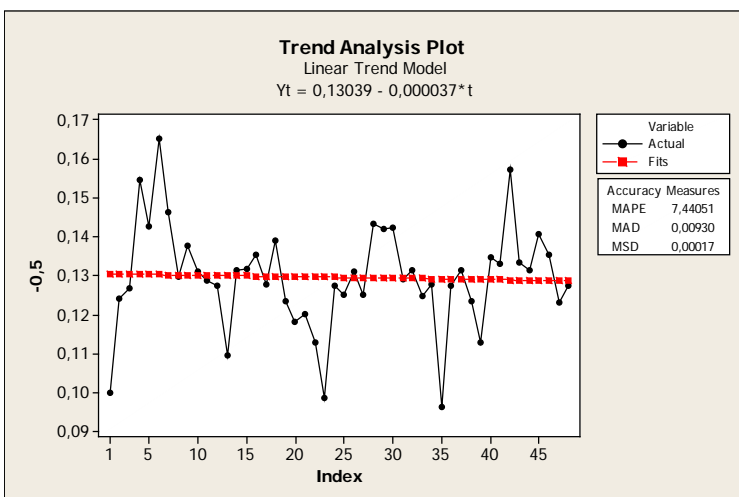
## Bilaga 1



Figur 1: Trendanalys för varugrupp ett, decembervärdena ersatta med medelvärdet för resterande serie.

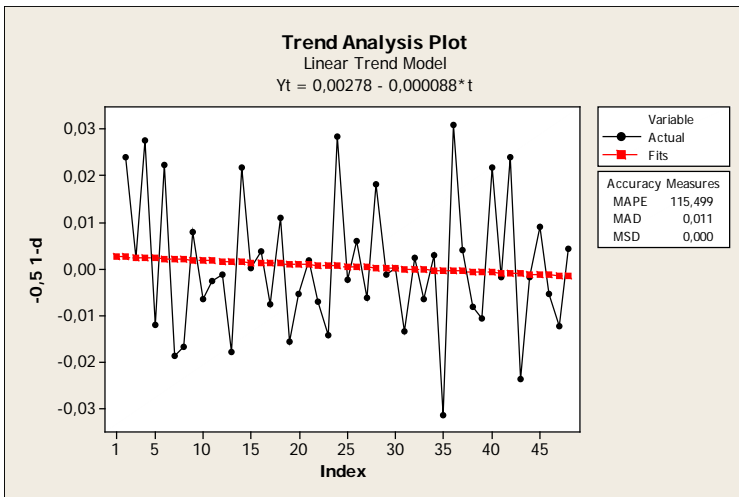


Figur 2: Box-Cox plott över varugrupp ett efter att decembervärdena är utbytta.

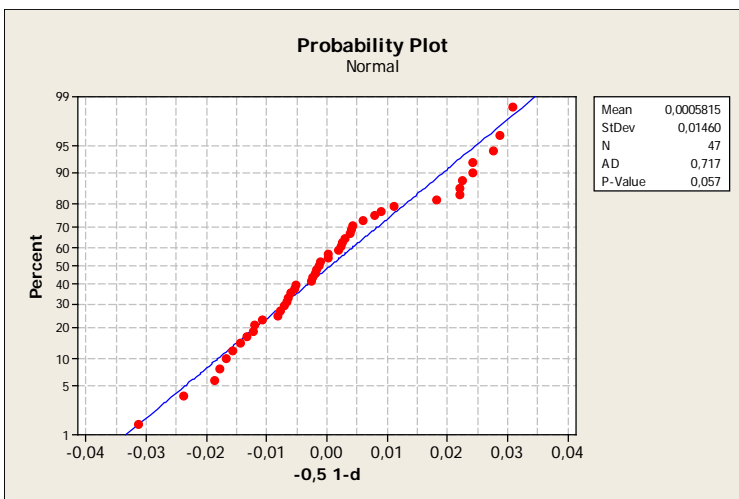


Figur 3: Trendanalys för varugrupp ett efter -0,5-transformation

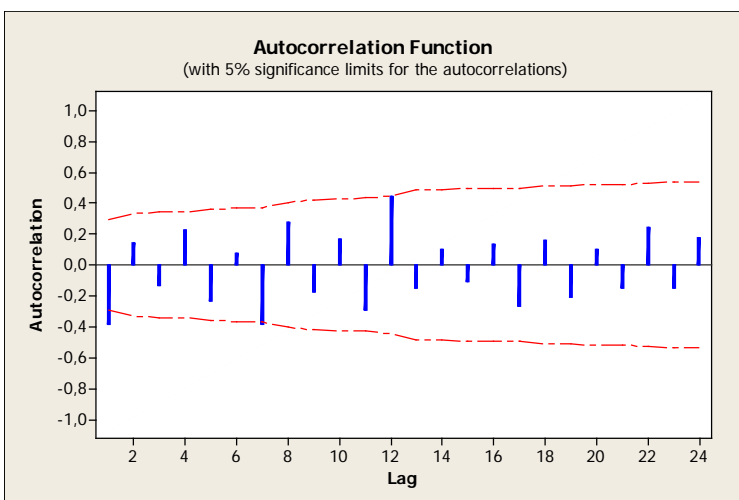




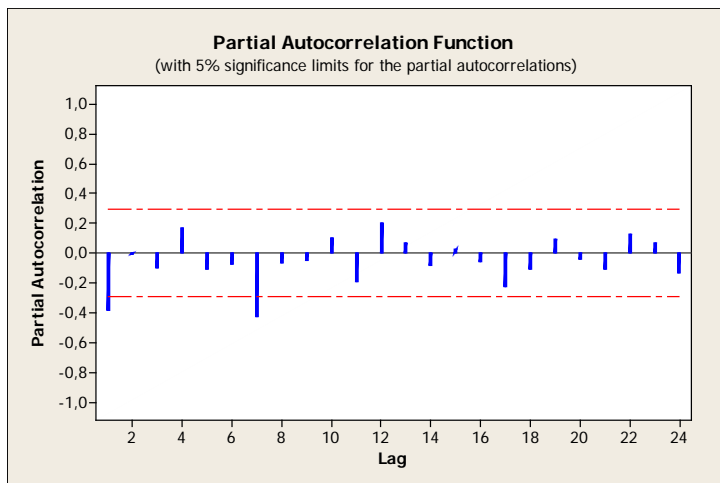
Figur 4: Trendanalys för varugrupp ett efter -0,5-transformering och första differentiering.



Figur 5: Normalitetsplott för varugrupp ett efter -0,5-transformering och första differentiering.



Figur 6: Autokorrelationsfunktion för varugrupp ett efter -0,5-transformering och första differentiering.



Figur 7: Partiell autokorrelationsfunktion för varugrupp ett efter -0,5-transformering och första differentiering.

### Utskrift 1: Minitabs utskrift över slutliga modellen, MA(1)\*SMA(1)<sub>12</sub>

Final Estimates of Parameters

Type		Coef	SE Coef	T	P
MA	1	0,3081	0,1457	2,11	0,040
SMA	12	-0,5275	0,1712	-3,08	0,004

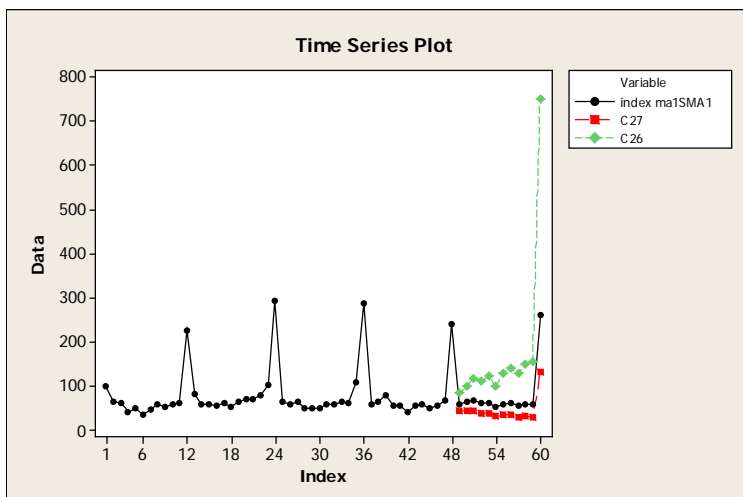
Differencing: 1 regular difference

Number of observations: Original series 48, after differencing 47

Residuals: SS = 0,00509040 (backforecasts excluded)  
MS = 0,00011312 DF = 45

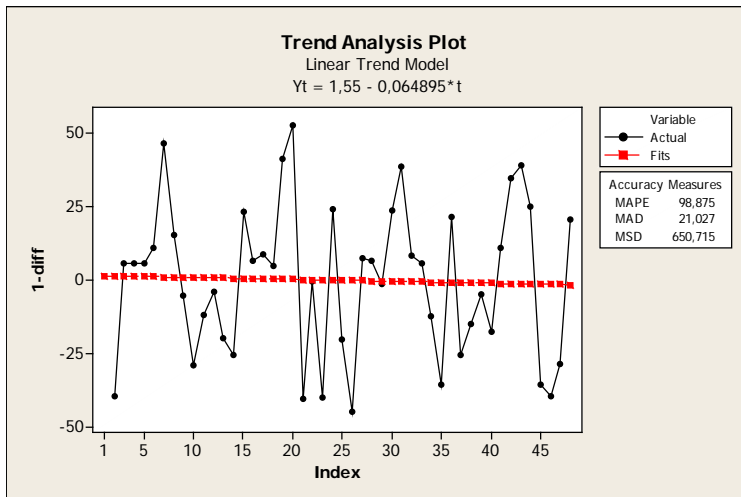
Modified Box-Pierce (Ljung-Box) Chi-Square statistic

Lag	12	24	36	48
Chi-Square	16,0	25,3	40,1	*
DF	10	22	34	*
P-Value	0,099	0,283	0,219	*

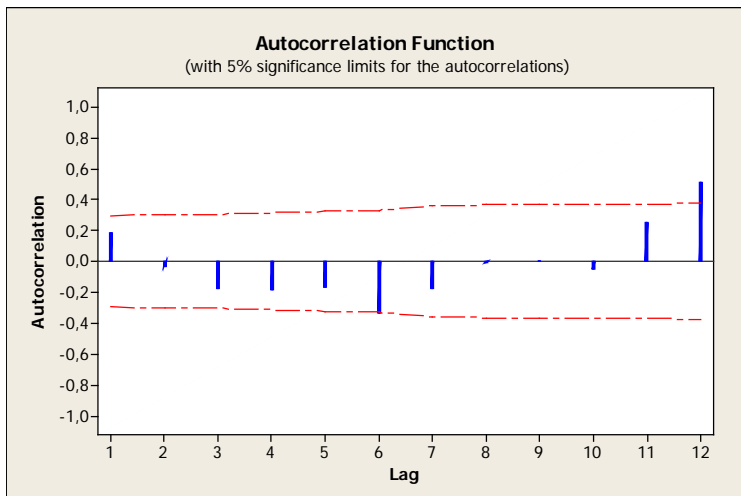


Figur 8: Prognosen med konfidensintervall för varugrupp ett.

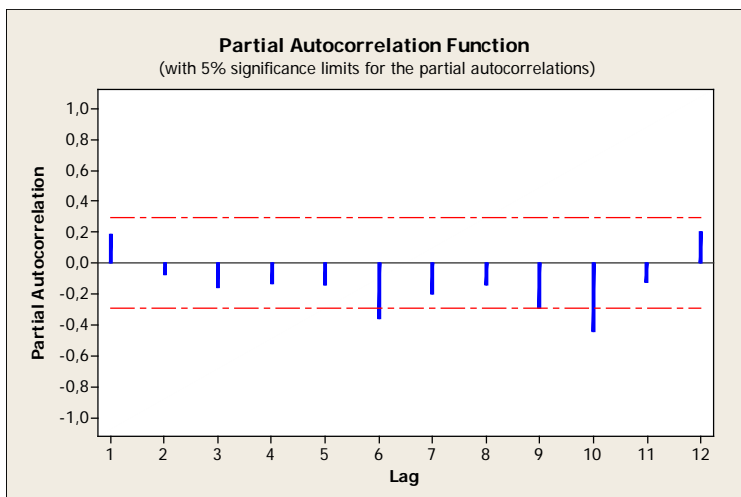
## Bilaga 2



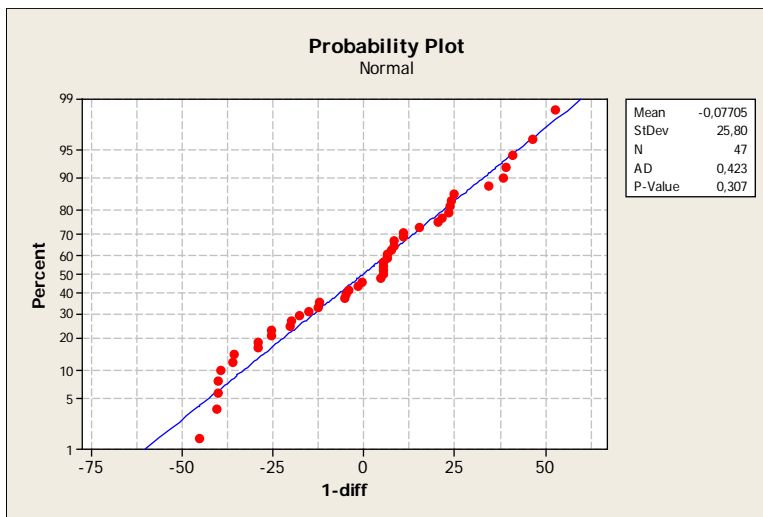
Figur 1: Trendanalys för varugrupp två efter första differentiering.



Figur 2: Autokorrelationsfunktion för varugrupp två efter första differentiering



Figur 3: Partiell autokorrelationsfunktion för varugrupp två efter första differentiering



Figur 4: Normalitetsplott för varugrupp två efter första differentiering

### Utskrift 1: Minitabs utskrift över slutliga modellen, SMA(1)<sub>12</sub>

Final Estimates of Parameters

Type	Coef	SE Coef	T	P
SMA 12	-0,7523	0,1851	-4,06	0,000

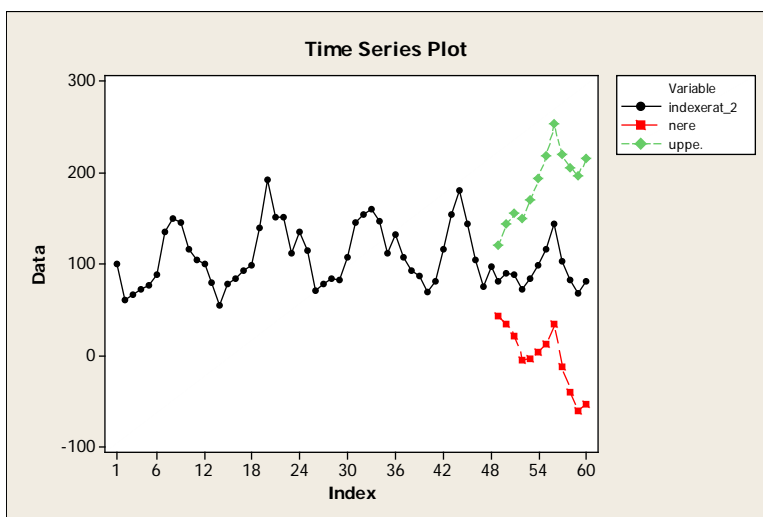
Differencing: 1 regular difference

Number of observations: Original series 48, after differencing 47

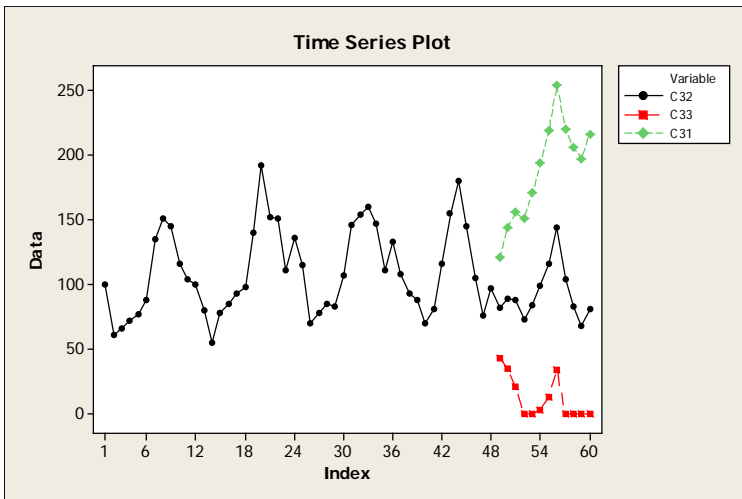
Residuals: SS = 18016,1 (backforecasts excluded)  
MS = 391,7 DF = 46

Lag	Chi-Square	DF	P-Value
12	12,0	11	0,366
24	25,5	23	0,323
36	35,2	35	0,458
48	*	*	*

Modified Box-Pierce (Ljung-Box) Chi-Square statistic

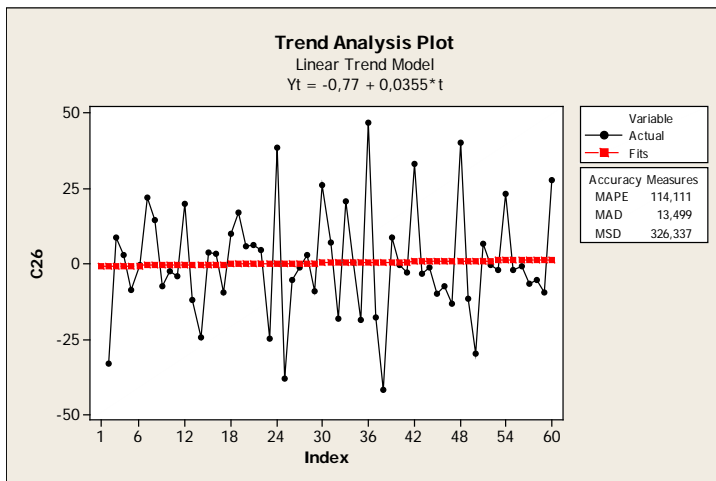


Figur 5: Prognosen med konfidensintervall för varugrupp två

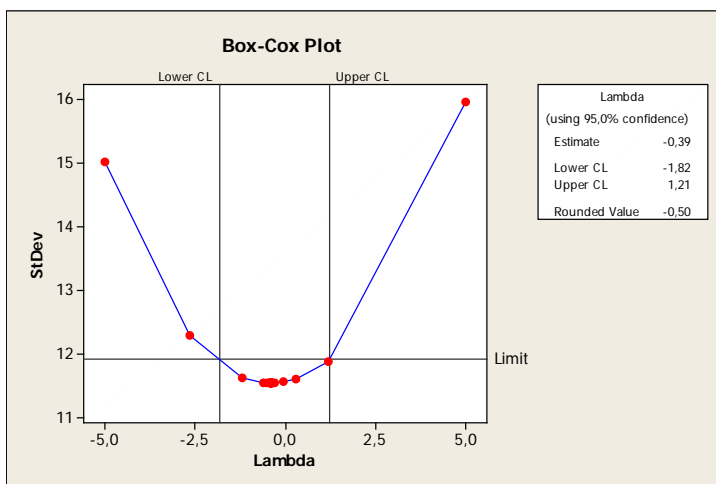


Figur 6: Prognosen med konfidensintervall för varugrupp två med antagande om endast positiv försäljning

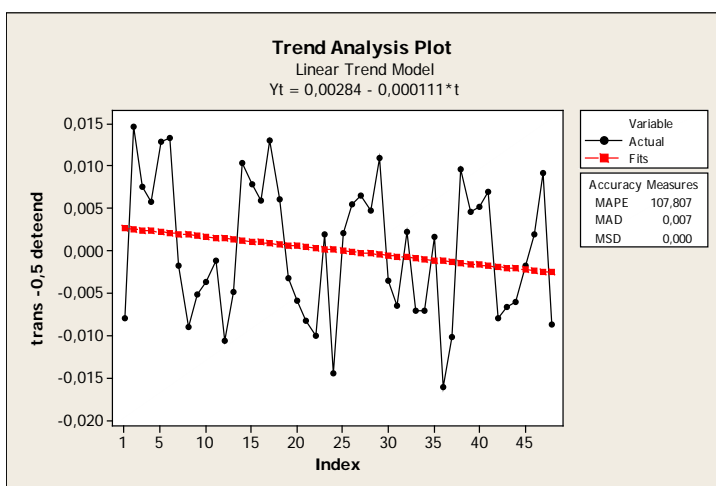
### Bilaga 3



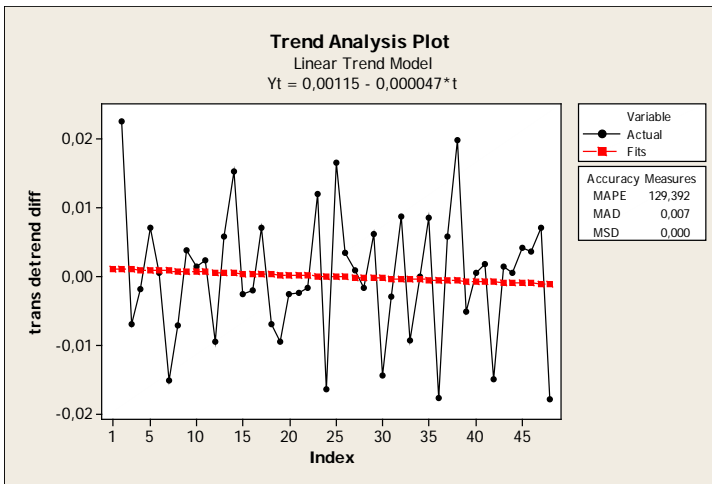
Figur 1: Trendanalys för varugrupp tre efter första differentiering.



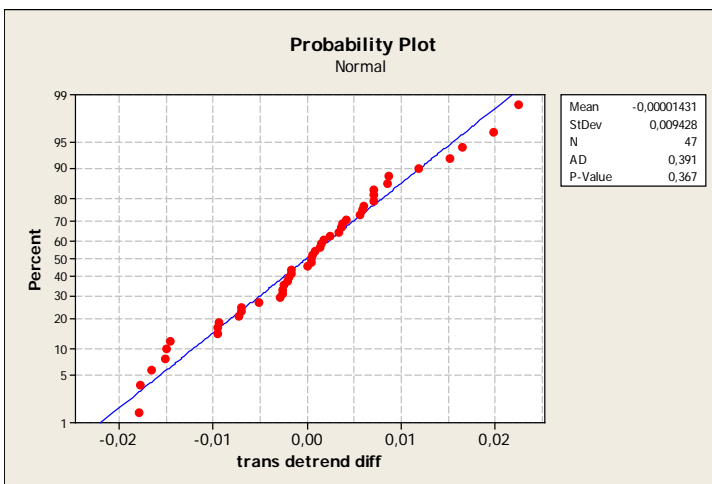
Figur 2: Box-Cox test av ojusterad serie



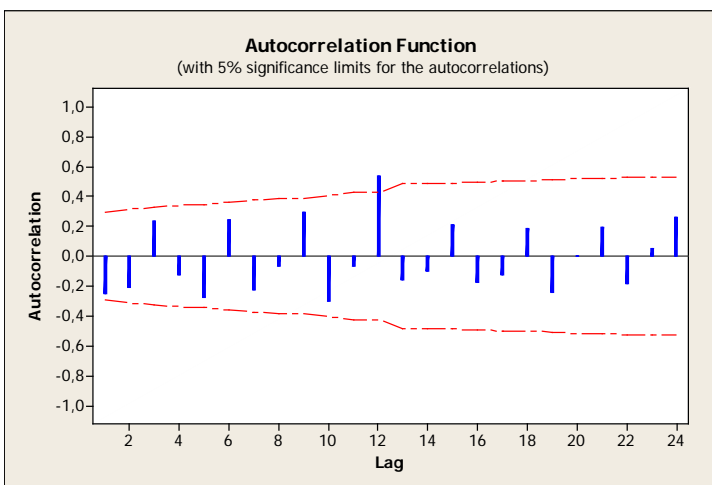
Figur 3: Trendanalys för varugrupp tre efter avtrendning och -0,5-transformering.



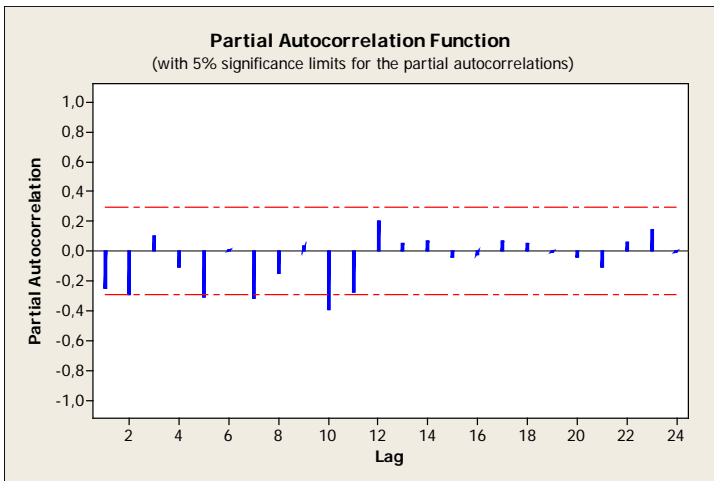
Figur 4: Trendanalys för varugrupp tre efter avtrendning, -0,5-transformering och första differentiering.



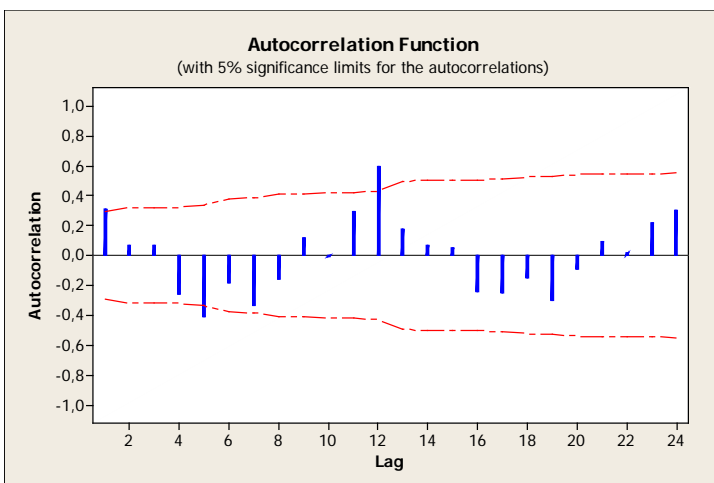
Figur 5: Normalitetsplott för varugrupp tre efter avtrendning, -0,5-transformering och första differentiering.



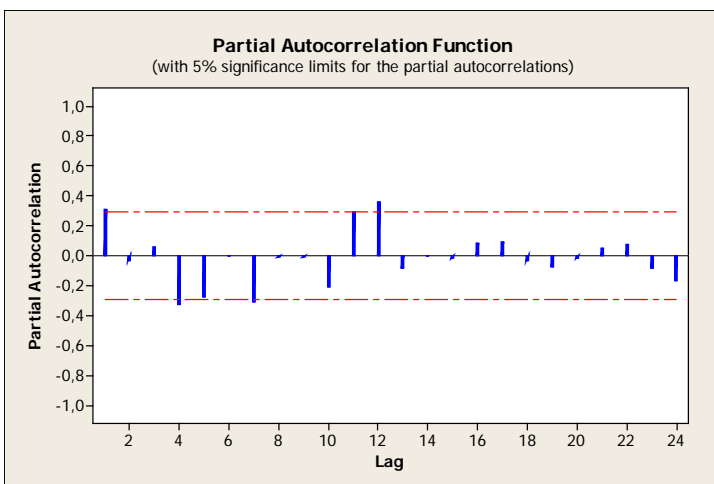
Figur 6: Autokorrelationsfunktion för varugrupp tre efter avtrendning, -0.5-transformering och första differentiering.



Figur 7: Partiell autokorrelationsfunktion för varugrupp tre efter avtrendning, -0.5-transformering och första differentiering.

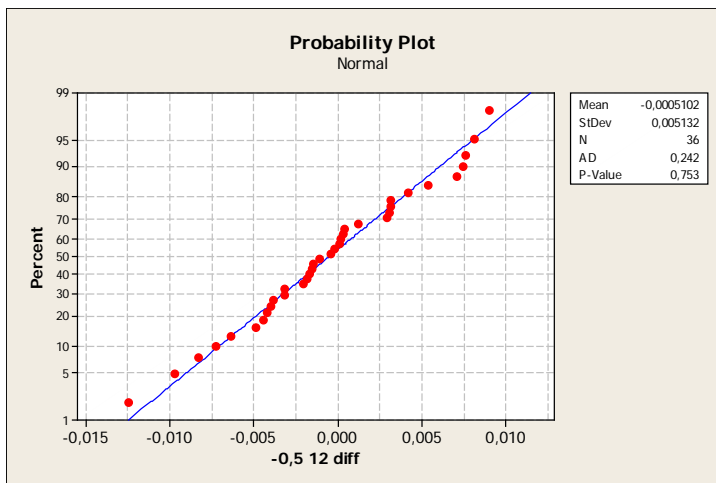


Figur 8: Autokorrelationsfunktion för varugrupp tre efter avtrendning och -0,5-transformering.

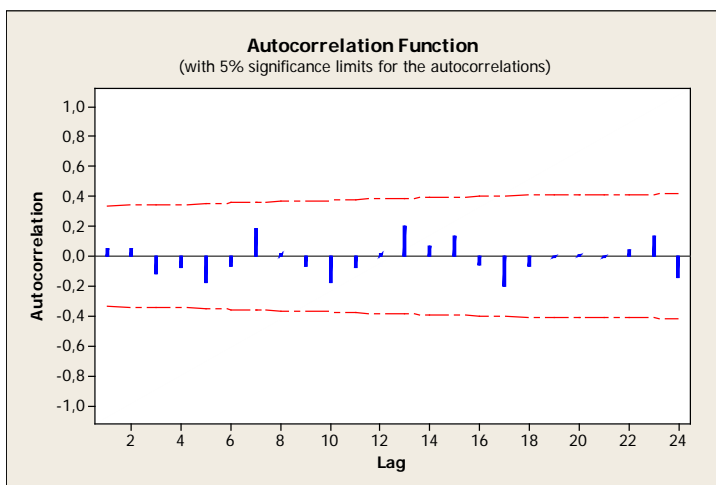


Figur 9: Partiell autokorrelationsfunktion för varugrupp tre efter avtrendning och -0,5-transformering.

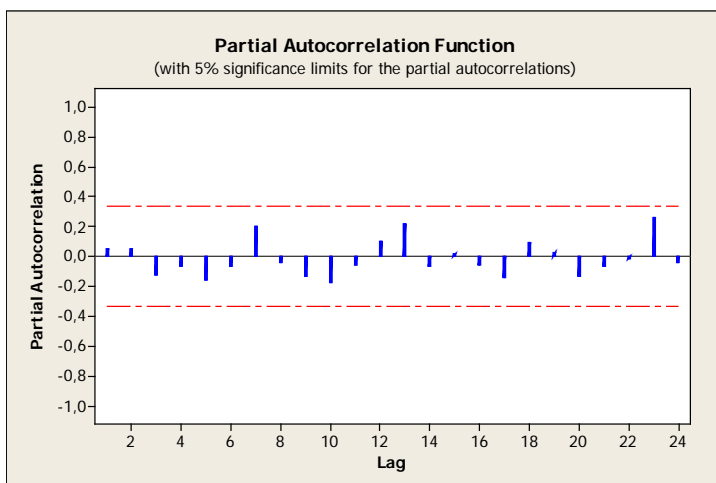




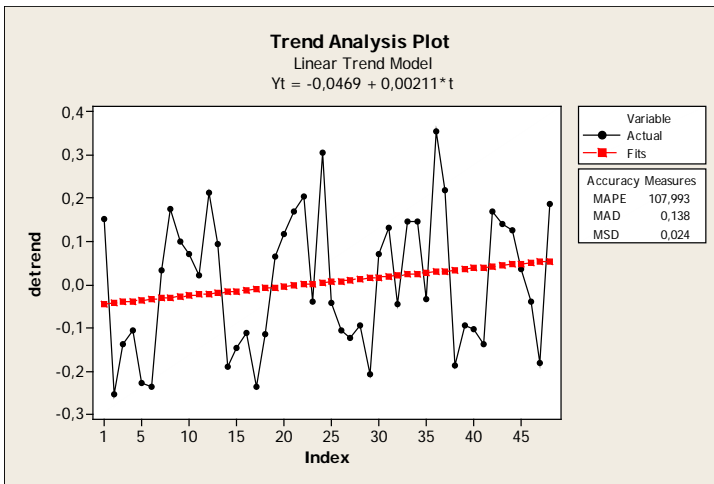
Figur 10: Normalitetsplott för varugrupp tre efter avtrendning, -0,5-transformering och tolfte differentiering.



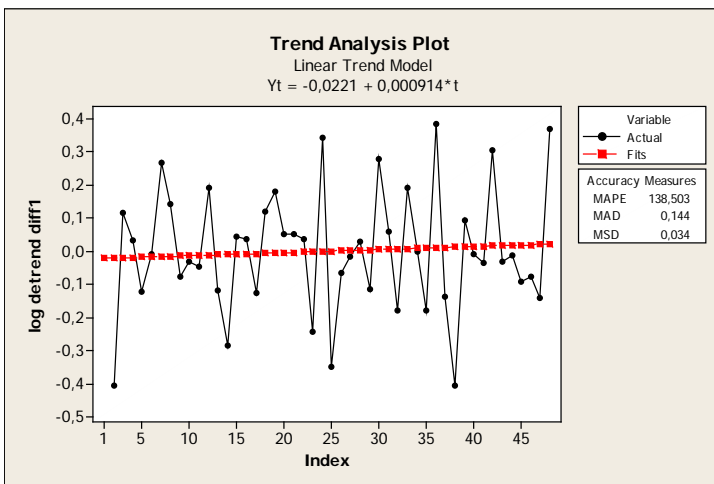
Figur 11: Autokorrelationsfunktion för varugrupp tre efter avtrendning, -0,5-transformering och tolfte differentiering.



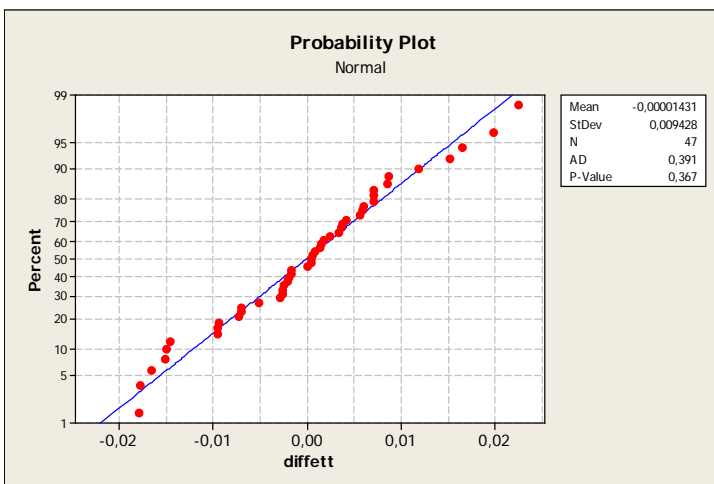
Figur 12: Partiell autokorrelationsfunktion för varugrupp tre efter avtrendning, -0,5-transformering och tolfte differentiering.



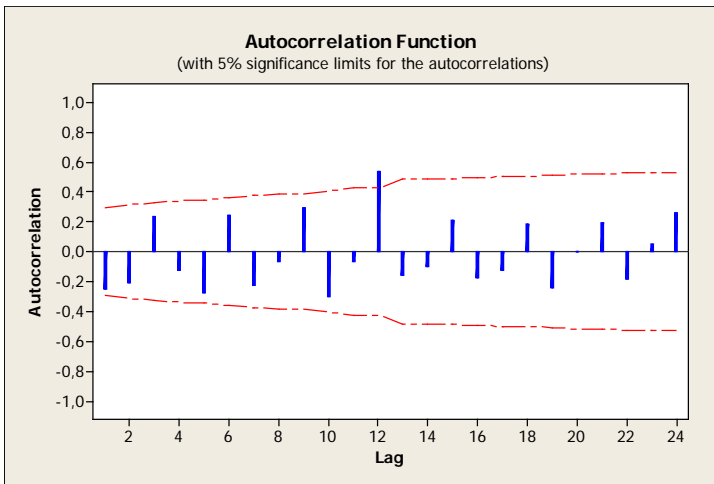
Figur 13: Trendanalys för varugrupp tre efter avtrendning och log-transformering.



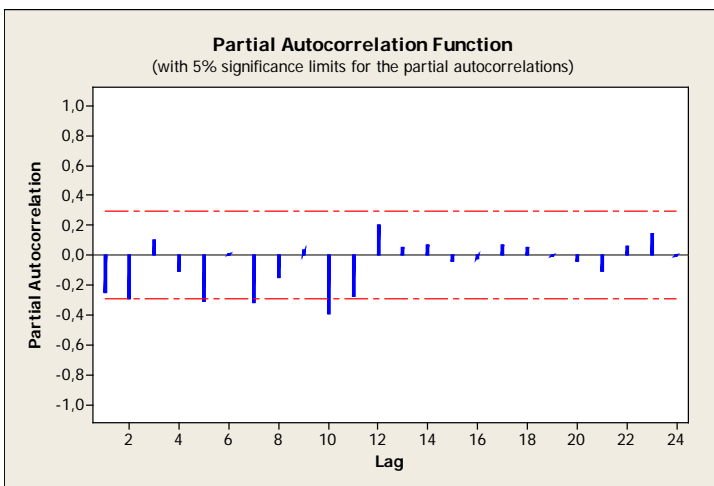
Figur 14: Trendanalys för varugrupp tre efter avtrendning, log-transformering och första differentiering.



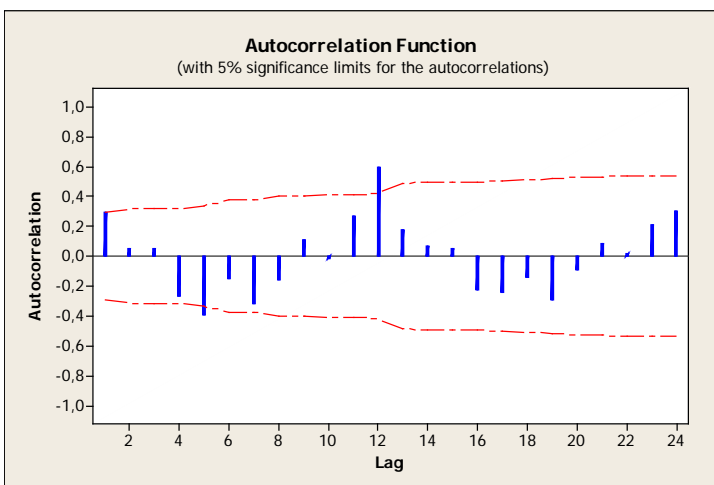
Figur 15: Normalitetsplott för varugrupp tre efter avtrendning, log-transformering och första differentiering.



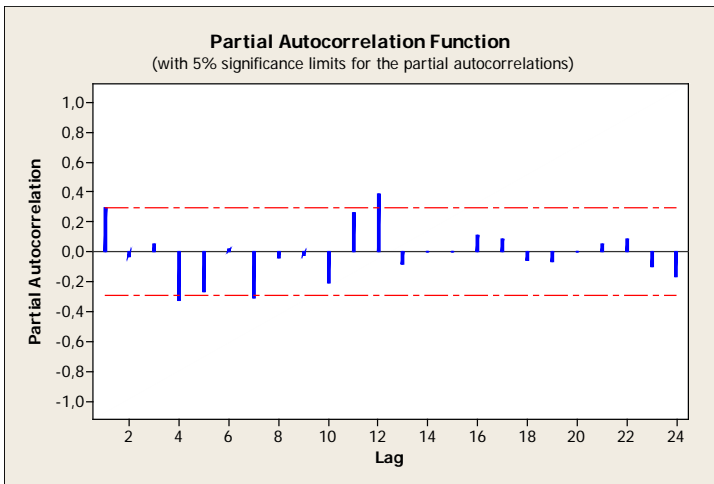
Figur 16: Autokorrelationsfunktion för varugrupp tre efter avtrendning, log-transformering och första differentiering.



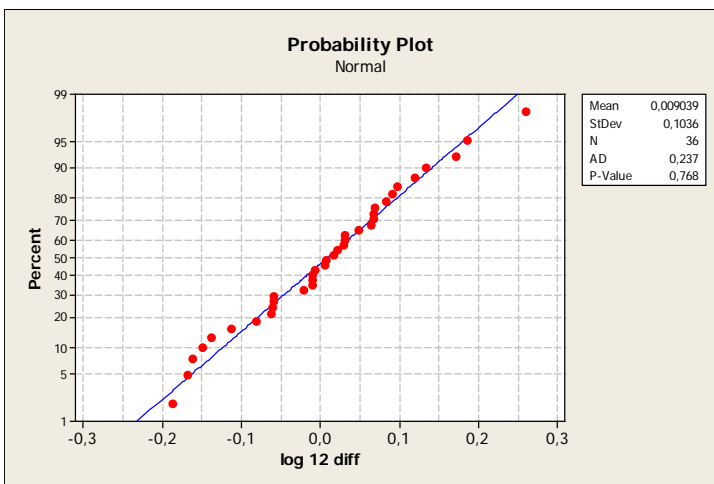
Figur 17: Partiell autokorrelationsfunktion för varugrupp tre efter avtrendning, log-transformering och första differentiering.



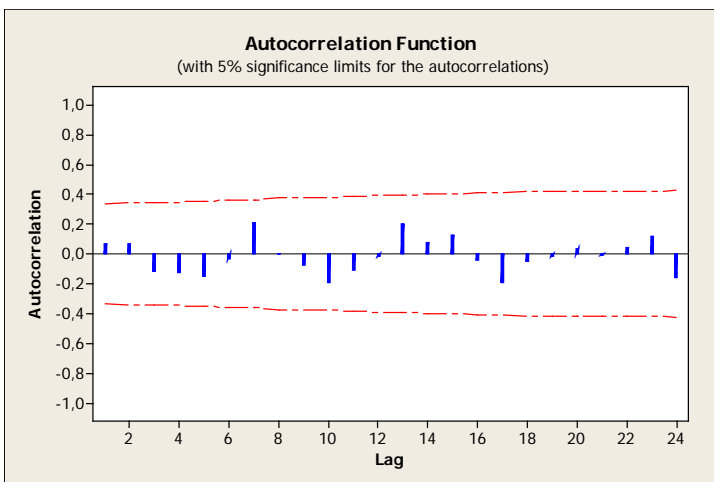
Figur 18: Autokorrelationsfunktion för varugrupp tre efter avtrendning och log-transformering.



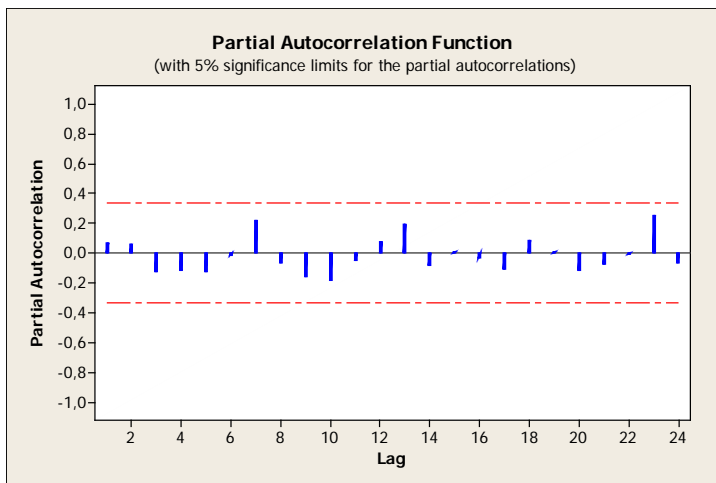
Figur 19: Partiell autokorrelationsfunktion för varugrupp tre efter avtrendning och log-transformering.



Figur 20: Normalitetsplott för varugrupp tre efter avtrendning, log-transformering och tolfte differentiering.



Figur 21: Autokorrelationsfunktion för varugrupp tre efter avtrendning, log-transformering och tolfte differentiering.



Figur 22: Partiell autokorrelationsfunktion för varugrupp tre efter avtrendning, log-transformering och tolfte differentiering.

### Utskrift 1: Minitabs utskrift av SAR(1)<sub>12</sub> modell efter avtrendning och -0,5-transformering

Final Estimates of Parameters

Type	Coef	SE Coef	T	P
SAR 12	0,7369	0,1173	6,28	0,000

Number of observations: 48

Residuals: SS = 0,00162579 (backforecasts excluded)  
MS = 0,00003459 DF = 47

Modified Box-Pierce (Ljung-Box) Chi-Square statistic

Lag	12	24	36	48
Chi-Square	8,6	17,8	25,2	*
DF	11	23	35	*
P-Value	0,656	0,767	0,888	*

### Utskrift 2: Minitabs utskrift av SAR(1)<sub>12</sub> modell efter avtrendning och log-transformering

Final Estimates of Parameters

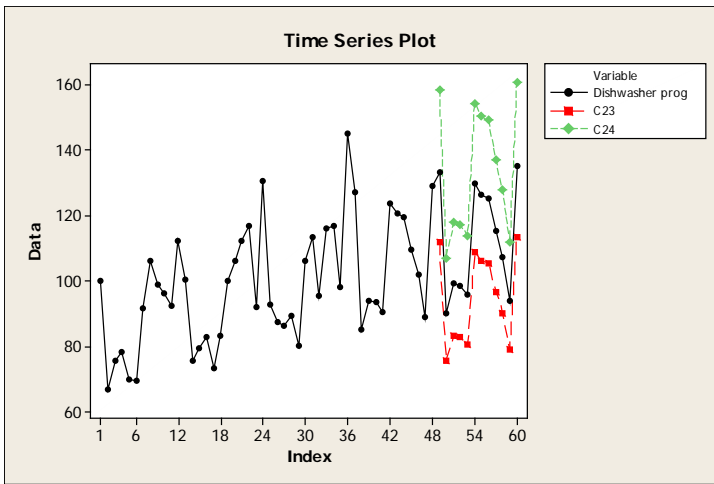
Type	Coef	SE Coef	T	P
SAR 12	0,9742	0,0800	12,18	0,000

Number of observations: 48

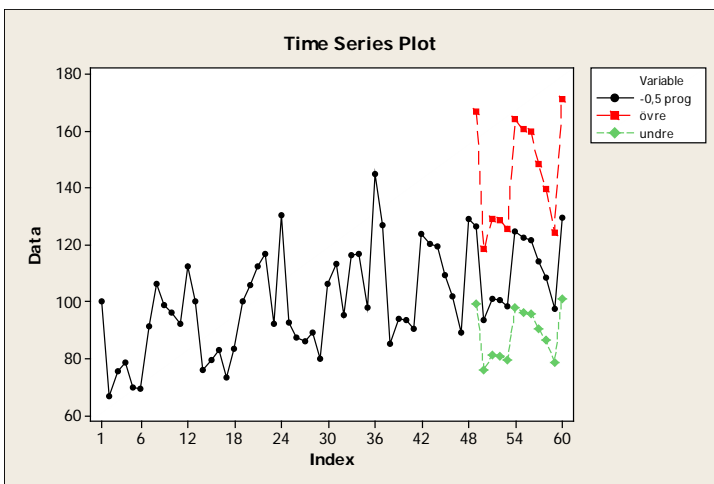
Residuals: SS = 0,368885 (backforecasts excluded)  
MS = 0,007849 DF = 47

Modified Box-Pierce (Ljung-Box) Chi-Square statistic

Lag	12	24	36	48
Chi-Square	9,9	21,3	22,6	*
DF	11	23	35	*
P-Value	0,542	0,563	0,948	*

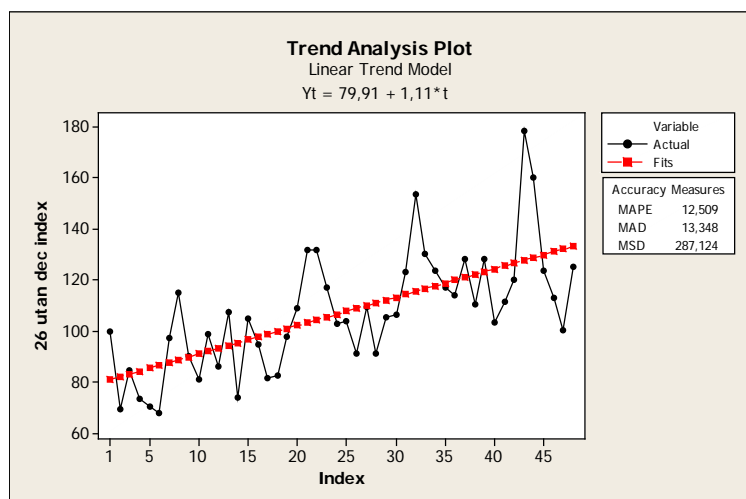


Figur 23: Serien samt prognos och konfidensintervall för log-transformerad serie

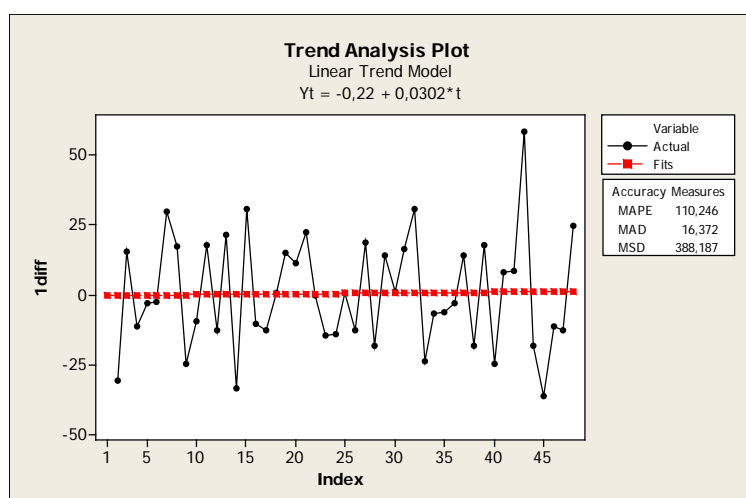


Figur 24: Serien samt prognos och konfidensintervall för -0,5-transformerad serie

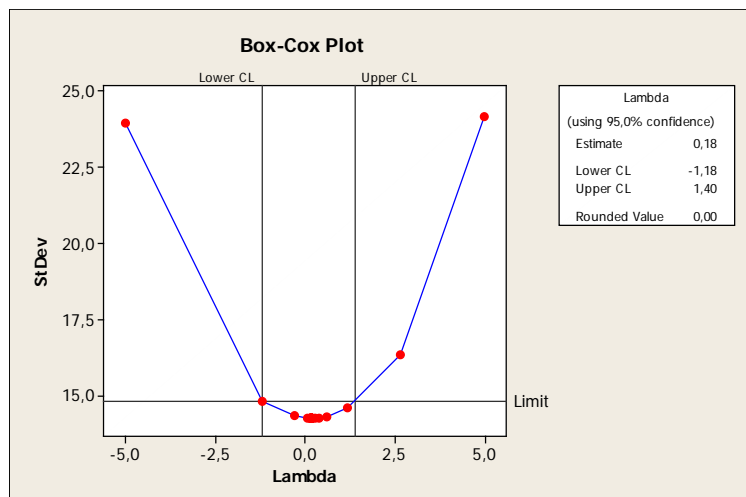
## Bilaga 4



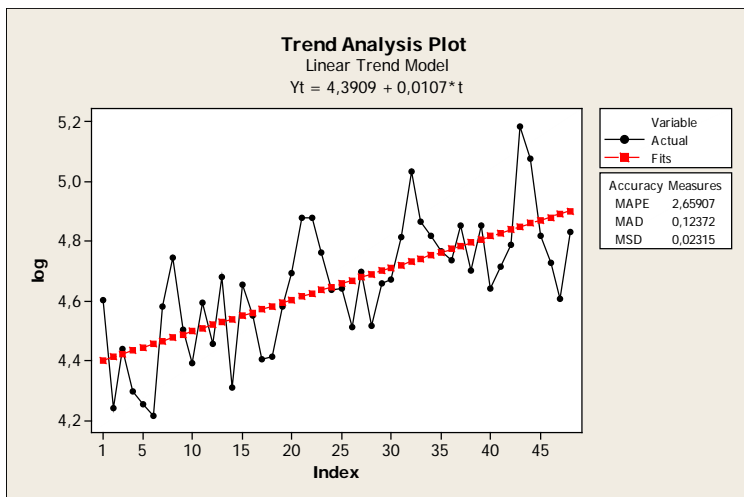
Figur 1: Trendanalys för varugrupp fyra, decembervärdena ersatta med medelvärdet för varje års resterande serie.



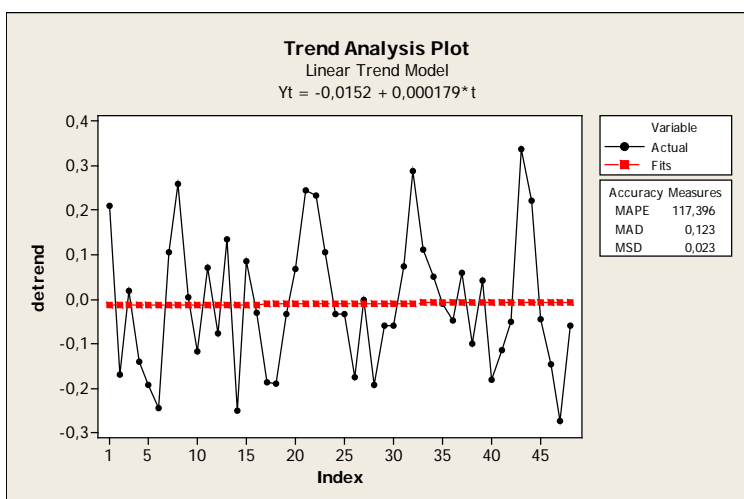
Figur 2: Trendanalys för varugrupp fyra efter första differentiering.



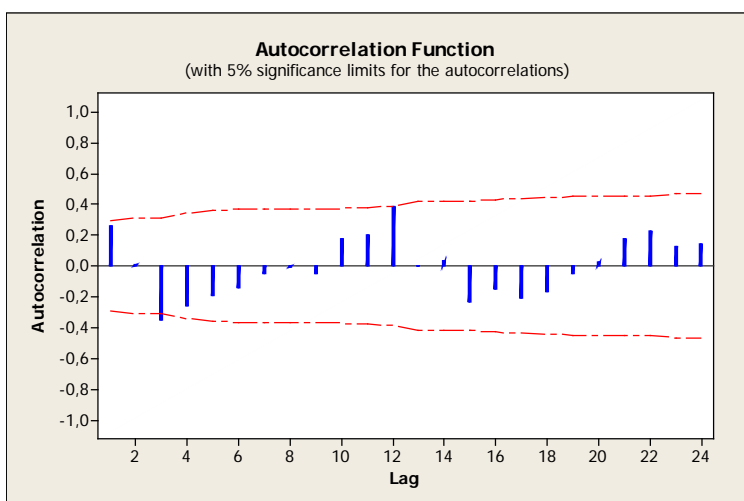
Figur 3: Box-Cox test av ojusterad serie.



Figur 4: Trendanalys av varugrupp fyra efter log-transformering.

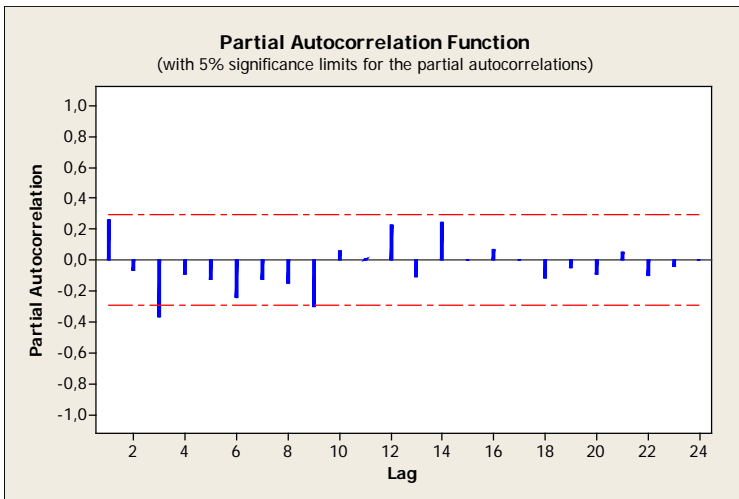


Figur 5: Trendanalys av varugrupp fyra efter avtrendning och log-transformering.

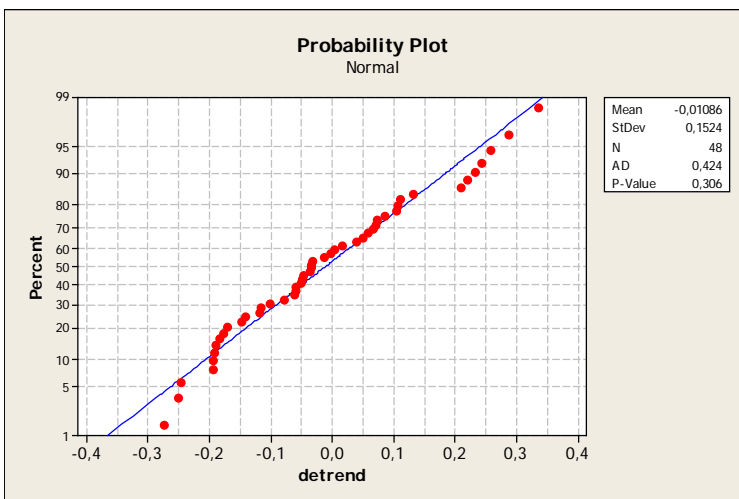


Figur 6: Autokorrelationsfunktion för varugrupp fyra efter avtrendning och log-transformering.

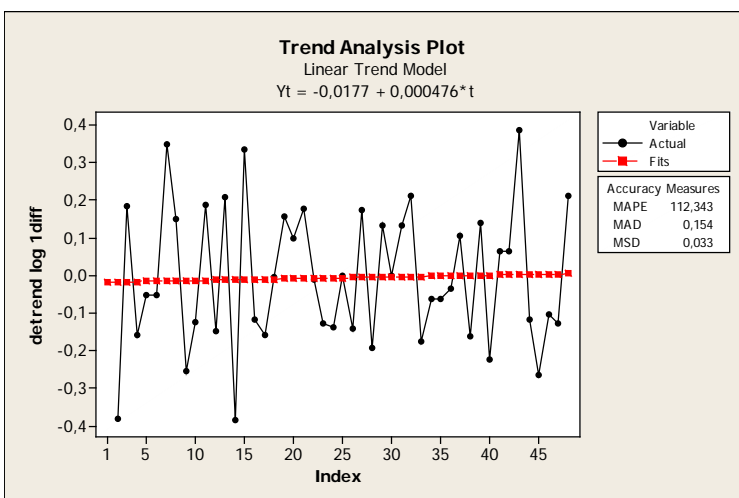




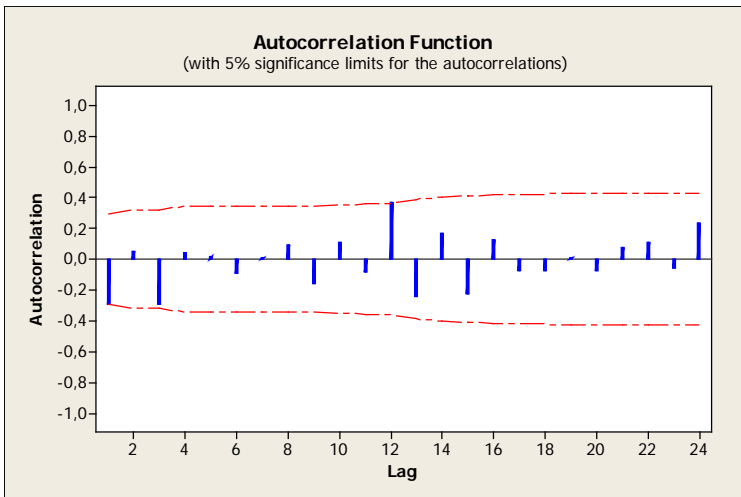
Figur 7: Partiell autokorrelationsfunktion för varugrupp fyra efter avtrendning och log-transformering.



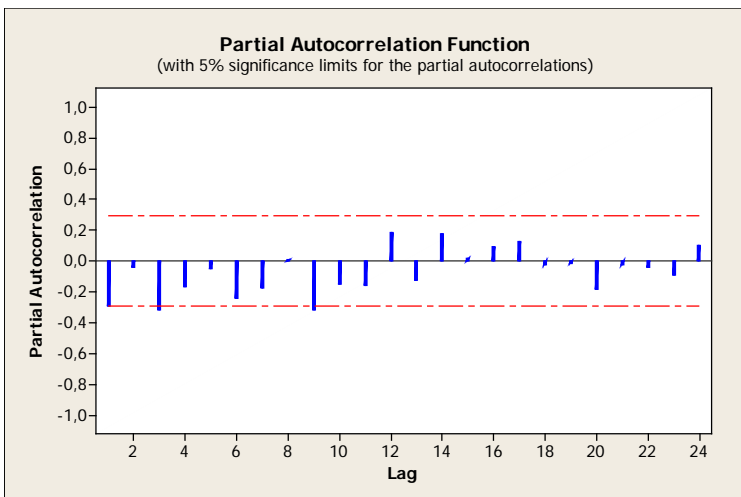
Figur 8: Normalitetsplott för varugrupp fyra efter avtrendning och log-transformering.



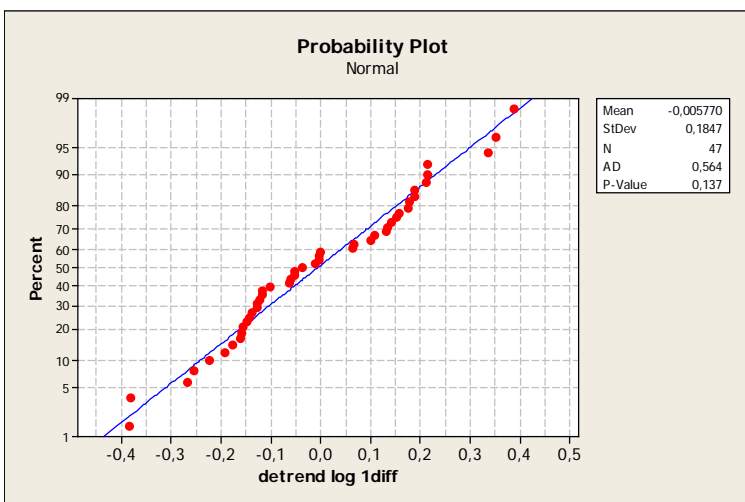
Figur 9: Trendanalys av varugrupp fyra efter avtrendning, log-transformering och första differentiering.



Figur 10: Autokorrelationsfunktion för varugrupp fyra efter avtrendning, log-transformering och första differentiering.



Figur 11: Partiell autokorrelationsfunktion för varugrupp fyra efter avtrendning, log-transformering och första differentiering.



Figur 12: Normalitetsplott för varugrupp fyra efter avtrendning, log-transformering och första differentiering.

Utskrift 1: Minitabs utskrift över SMA(1)<sub>12</sub> modell efter detrendning, log-transformering och första differentiering av serien.

Final Estimates of Parameters

Type	Coef	SE Coef	T	P
SMA 12	-0,7631	0,1547	-4,93	0,000

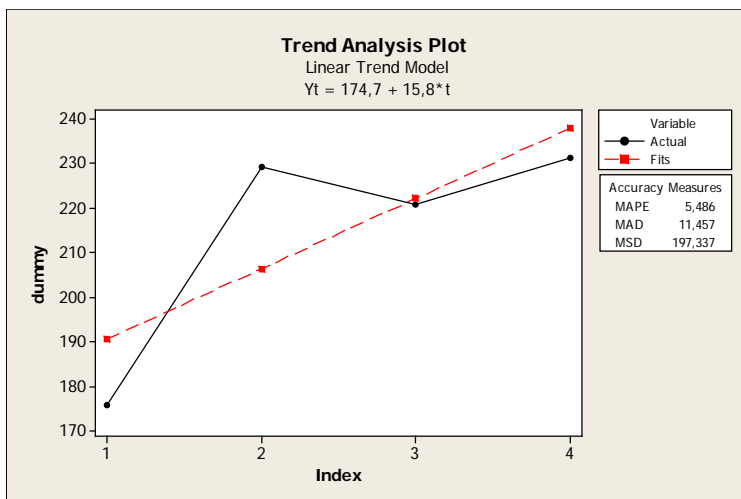
Differencing: 1 regular difference

Number of observations: Original series 48, after differencing 47

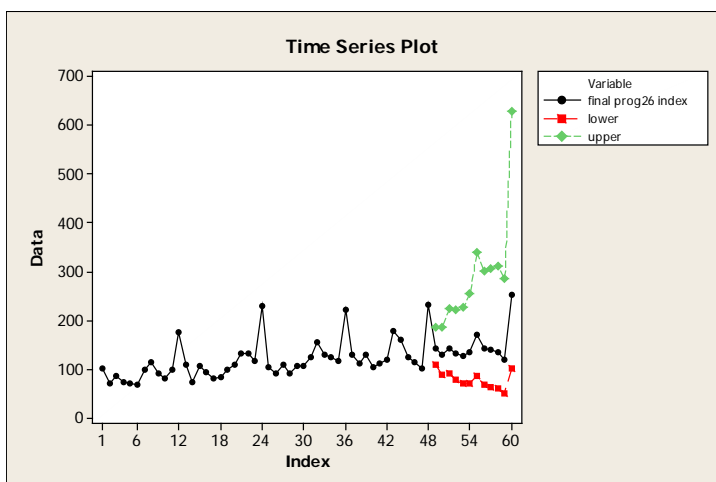
Residuals: SS = 0,823021 (backforecasts excluded)  
MS = 0,017892 DF = 46

Modified Box-Pierce (Ljung-Box) Chi-Square statistic

Lag	12	24	36	48
Chi-Square	10,4	25,4	39,1	*
DF	11	23	35	*
P-Value	0,493	0,330	0,291	*

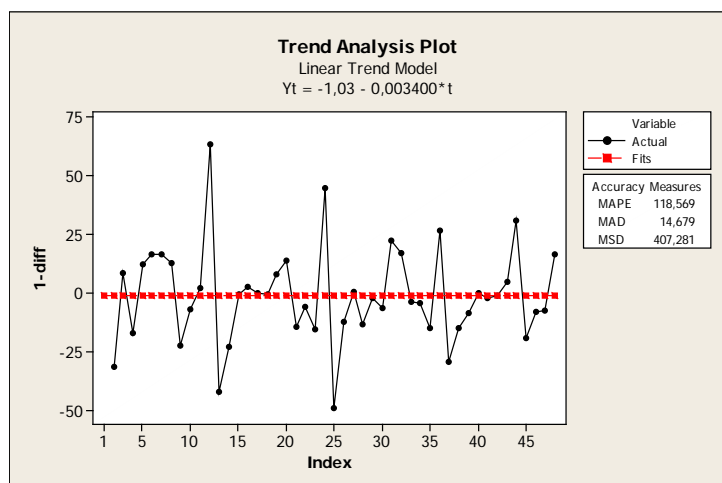


Figur 13: Trendanalys över decembervärdena.

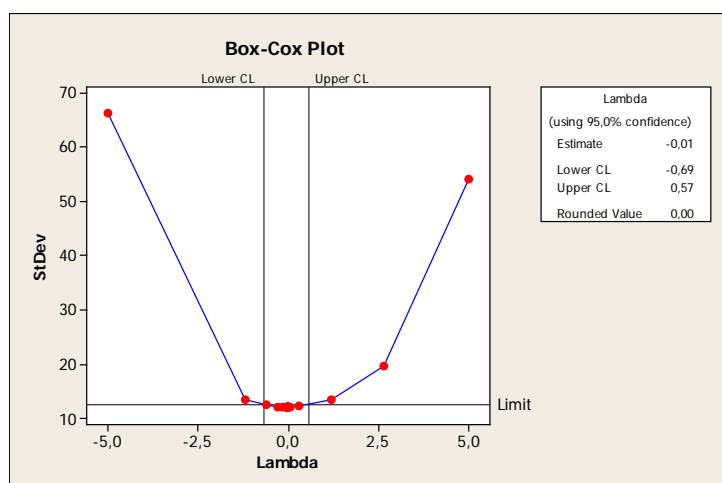


Figur 14: Serie med prognos och konfidensintervall för varugrupp 4.

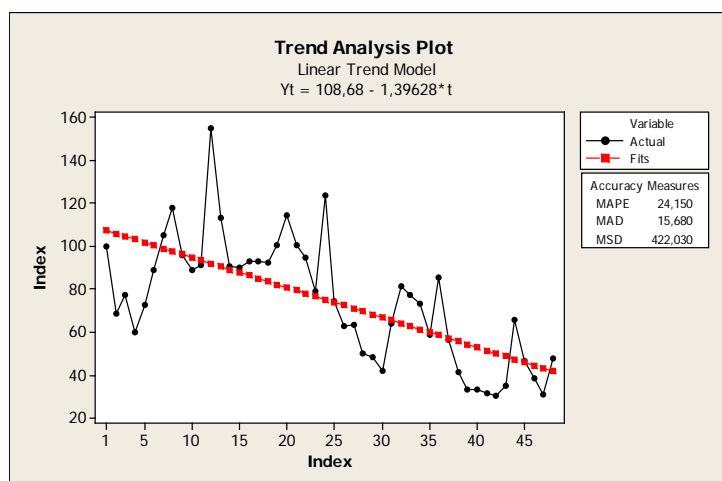
## Bilaga 5



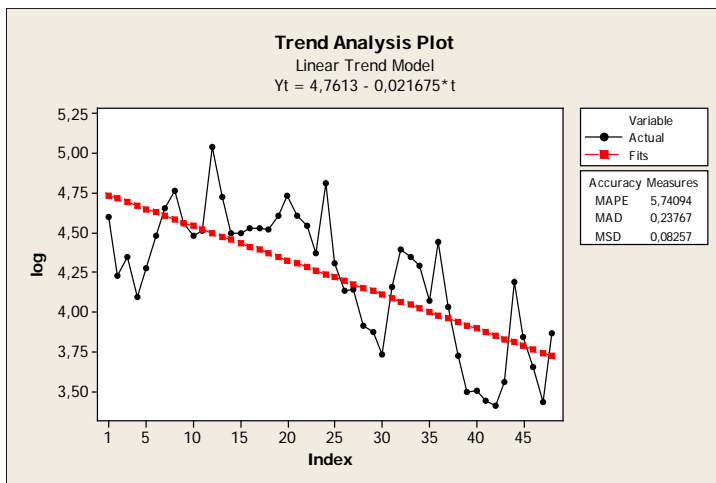
Figur 1: Trendanalys för varugrupp fem efter första differentiering.



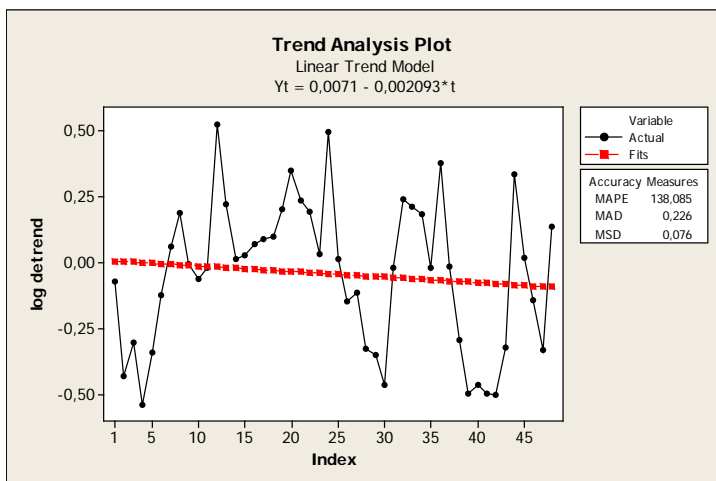
Figur 2: Box-Cox test av ojusterad serie



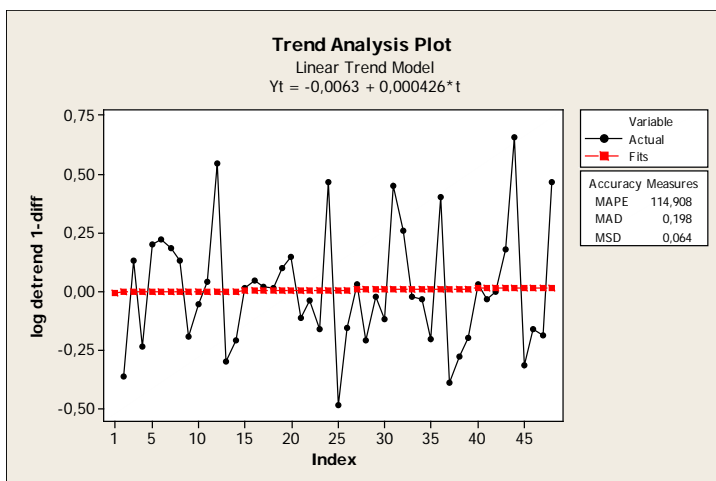
Figur 3: Trendanalys över ojusterad serie



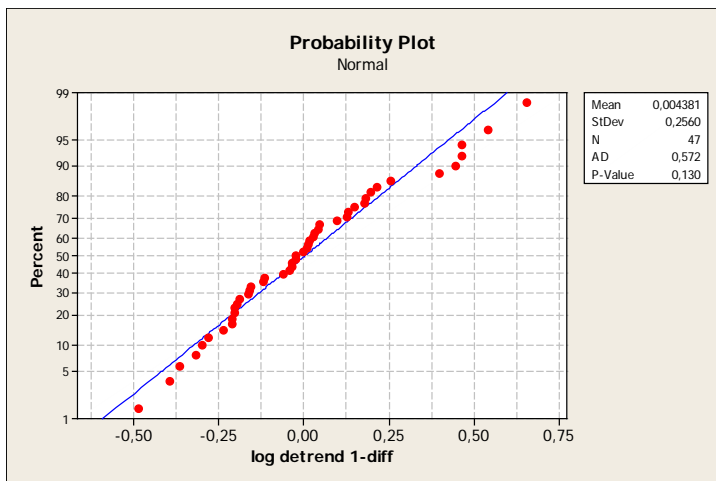
Figur 4: Trendanalys för varugrupp fem efter log-transformering.



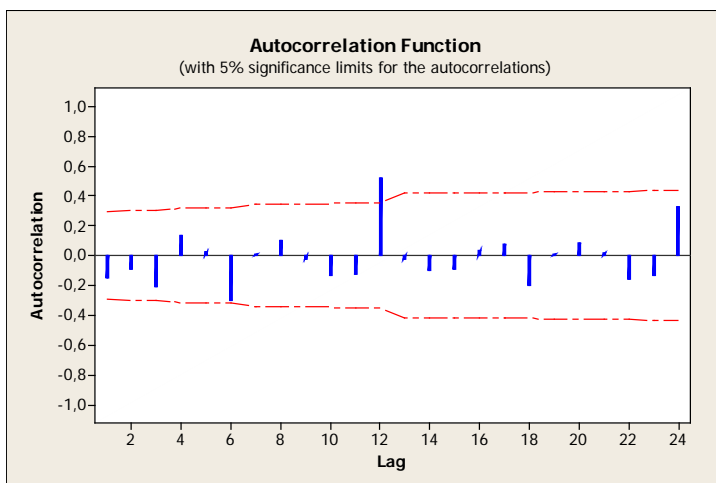
Figur 5: Trendanalys för varugrupp fem efter avtrendning och log-transformering.



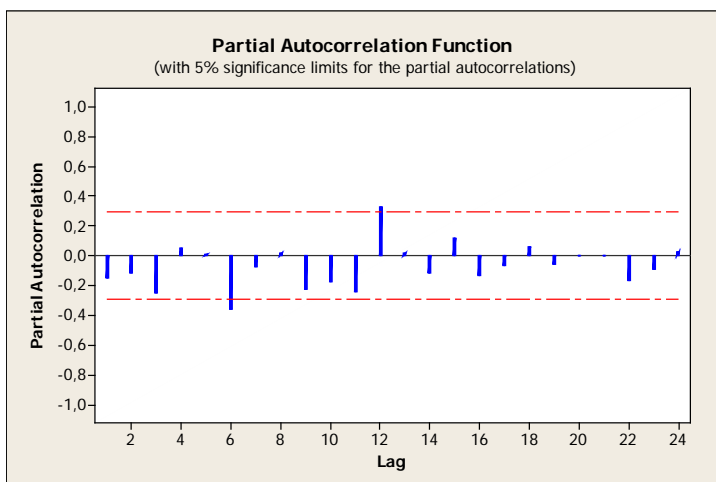
Figur 6: Trendanalys för varugrupp fem efter avtrendning, log-transformering och första differentiering.



Figur 7: Normalitetsplott över varugrupp fem efter avtrendning, log-transformering och första differentiering.



Figur 8: Autokorrelationsfunktion för varugrupp fem efter avtrendning, log-transformering och första differentiering.



Figur 9: Partiell autokorrelationsfunktion för varugrupp fem efter avtrendning, log-transformering och första differentiering.

# Utskrift 1: Minitabs utskrift över SAR(1)<sub>12</sub> modell efter detrendning, log-transformering och första differentiering av serien.

## Final Estimates of Parameters

Type	Coef	SE Coef	T	P
SAR 12	0,9860	0,0908	10,86	0,000

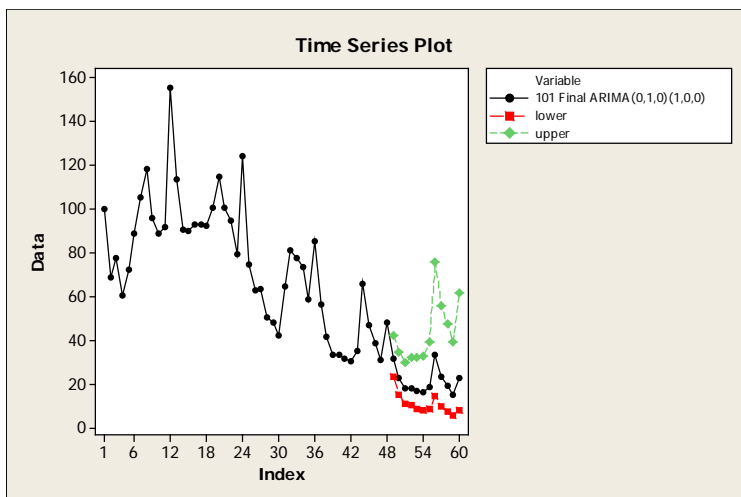
Differencing: 1 regular difference

Number of observations: Original series 48, after differencing 47

Residuals: SS = 1,00515 (backforecasts excluded)  
MS = 0,02185 DF = 46

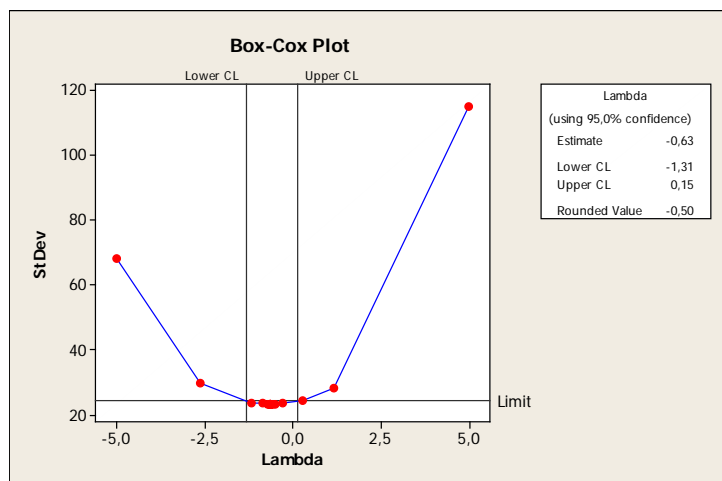
## Modified Box-Pierce (Ljung-Box) Chi-Square statistic

Lag	12	24	36	48
Chi-Square	14,4	26,1	39,5	*
DF	11	23	35	*
P-Value	0,209	0,295	0,275	*

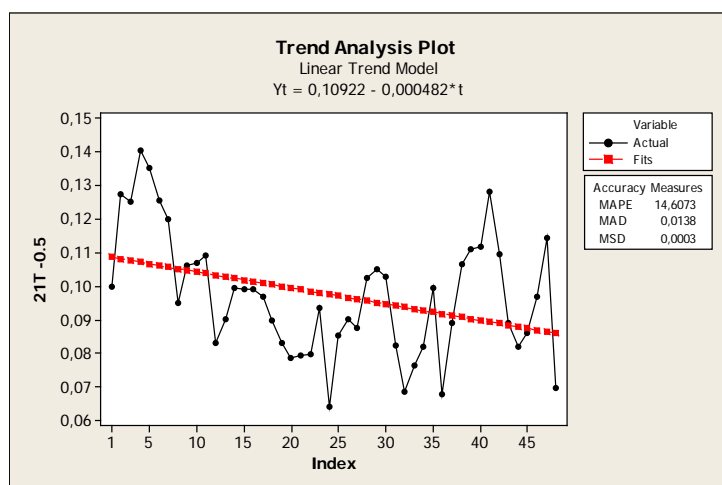


Figur 10: Serien med prognos och konfidensintervall för varugrupp 5

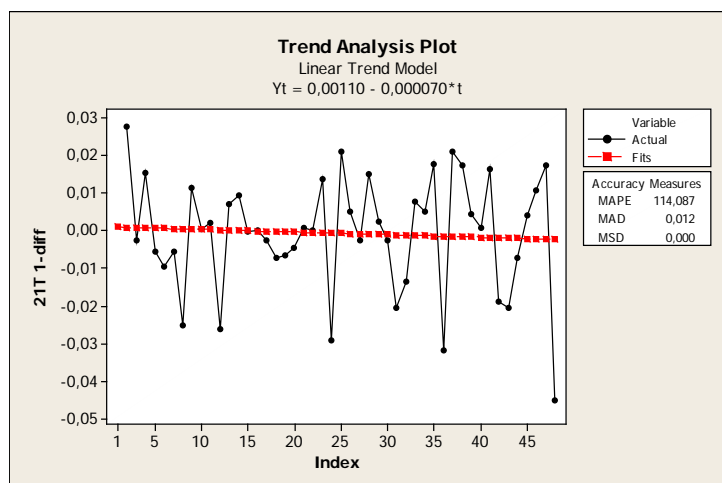
## Bilaga 6



Figur 1: Box-Cox test av ojusterad serie.

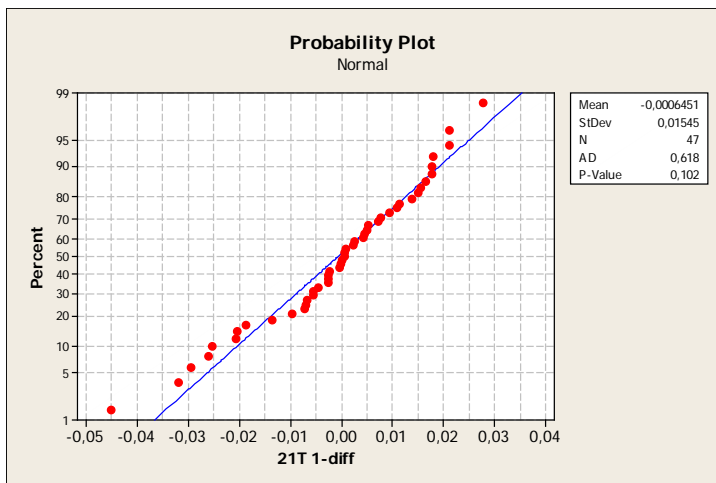


Figur 2: Trendanalys för varugrupp sex efter -0,5-transformering

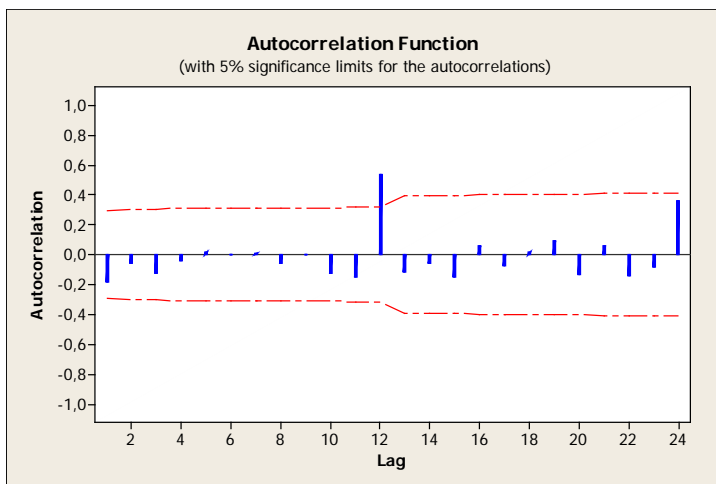


Figur 3: Trendanalys för varugrupp sex efter -0,5-transformering och första differentiering.

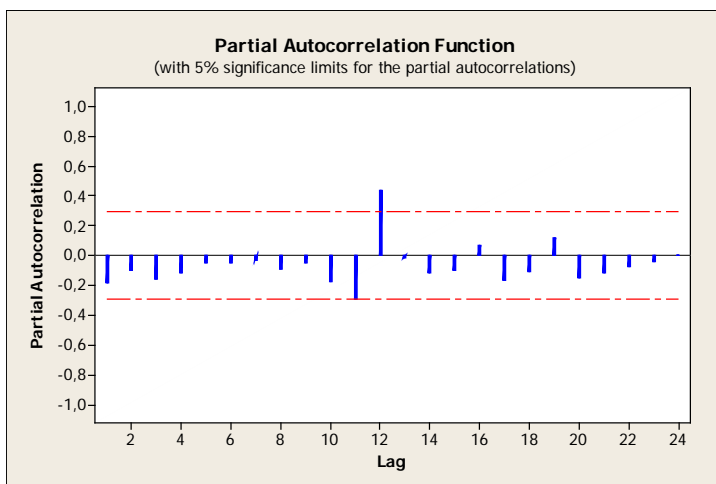




Figur 4: Normalitetsplott för varugrupp sex efter -0,5-transformering och första differentiering.



Figur 5: Autokorrelationsfunktion för varugrupp sex efter -0,5-transformering och första differentiering.



Figur 6: Partiell autokorrelationsfunktion för varugrupp sex efter -0,5-transformering och första differentiering.

Utskrift 1: Minitabs utskrift över SMA(1)<sub>12</sub> modell efter -0,5-transformering och första differentiering.

Final Estimates of Parameters

Type	Coef	SE Coef	T	P
SMA 12	-0,5501	0,1509	-3,65	0,001

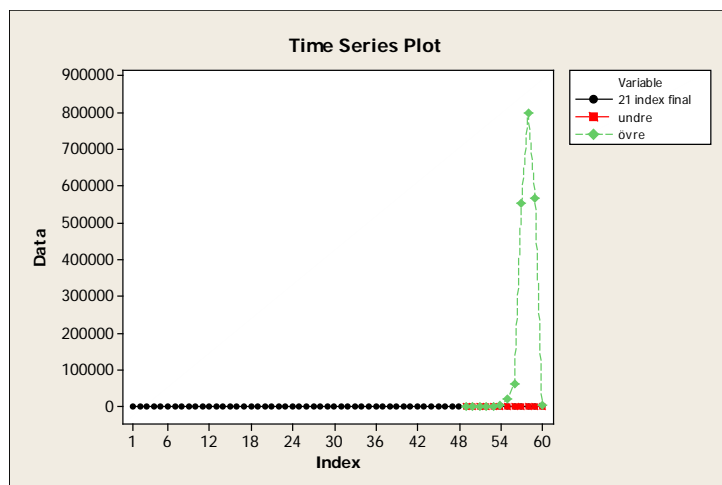
Differencing: 1 regular difference

Number of observations: Original series 48, after differencing 47

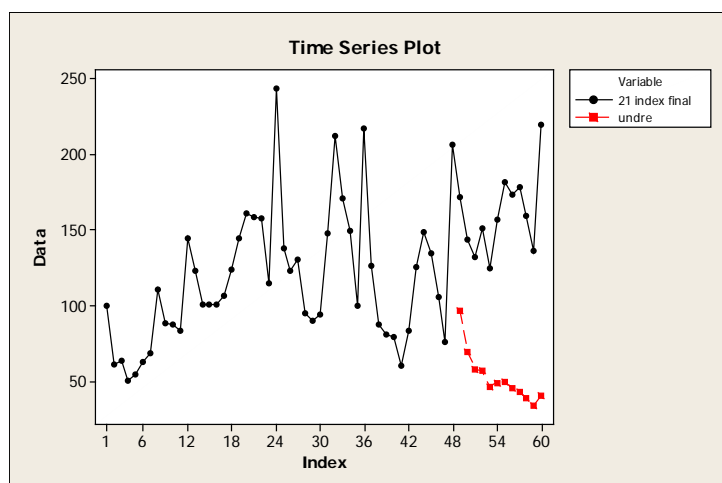
Residuals: SS = 0,00774732 (backforecasts excluded)  
MS = 0,00016842 DF = 46

Modified Box-Pierce (Ljung-Box) Chi-Square statistic

Lag	12	24	36	48
Chi-Square	6,0	29,5	37,9	*
DF	11	23	35	*
P-Value	0,875	0,165	0,338	*



Figur 7: Serien med prognos och konfidensintervall för varugrupp sex.



Figur 8: Serien med prognos och den undre intervallgränsen för varugrupp sex.