



**EKONOMIHÖGSKOLAN**  
Lunds universitet  
Statistiska institutionen

# **En analys av polismyndighetens prognosmodell LUPP – en hypotesprövning för Skåne län**

Thomas Arildsson

Uppsats i statistik  
10 poäng  
Nivå 61-80

Handledare: Mats Hagnell

Här vill jag passa på att rikta ett stort tack till polismyndigheten i Skåne och speciellt till Magdalena Lundqvist, som bistått med data och övrig information.

<u>Innehållsförteckning</u>	<u>Sidan</u>
1. Sammanfattning	3
2. Syfte	3
3. Kodning av brott	
3.1 Generellt om kodning av brott	4
3.2 Definition bilinbrott	4
4. Metod	
4.1 Introduktion till valda metoder	4
4.2 Genomförande och redovisning	5
4.3 Allmänt om prognoser	6
4.4 Prognosfel	7
4.5 Introduktion tidsserier	7
4.6 Tidsseriens innehåll	8
4.7 Att beakta vid skapandet av prognoser	9
4.8 Tidsserieuppdelning	
4.8.1 Inledning	9
4.8.2 Skattning av trendkomponent	10
4.9 Box-Jenkins	
4.9.1 Olika Box-Jenkins modeller	11
4.9.2 Identifikation av lämplig modell	12
4.9.3 Anpassning av Box-Jenkins modeller	14
4.9.4 Diagnostik	17
4.9.5 Box-Jenkins säsongsm modeller	19
4.9.6 Prognosmodellering utifrån Box-Jenkins modeller	20
4.10. LUPP	
4.10.1 Beskrivning av LUPP	21
4.10.2 LUPP i teorin	21
4.10.3 Prognosframställan via LUPP	22

4.11 Introduktion till statistiska tester	
4.11.1 Parametriska metoder	23
4.11.2 Icke-parametriska metoder	23
4.1.2 Formulering av hypoteser	
4.12.1 Undersökning av prognoskvalitén – samtliga polisområden	24
4.12.2 Undersökning av prognoskvalitén – enskilt polisområde	24
4.12.3 Undersökning av prognoskvalitén med avseende på LUPP	26
5. Redovisning av resultat	
5.1 Undersökning av prognoskvalitén – samtliga polisområden	26
5.2 Undersökning av prognoskvalitén – enskilt polisområde	27
5.3 Undersökning av prognoskvalitén med avseende på LUPP	27
6. Slutdiskussion	29
7. Summary in English	30

## Bilagor

Valda modeller - Tidsserieuppdelning

Valda modeller - ARIMA

Antal bilinbrott 1998-2004

Kruskal-Wallis test vid punkt 5.2

Kruskal-Wallis test vid punkt 5.3

Definition av polisområden

Källförteckning

## 1. Sammanfattning

Att försöka förutse framtiden är väsentligt för alla former av verksamheter, så även för den polisiära. Som i de flesta organisationer är resurserna begränsade, vilket gör att dessa på ett optimalt sätt måste prioriteras och fördelas. En viktig beståndsdel i detta allokeringsarbete är bland annat skapandet av prognoser. Att på ett tillförlitligt sätt kunna uppskatta och förutse den framtida brottsligheten, skapar ett viktigt underlag för att nå en effektiv resursallokering. Arbetet med att skapa prognoser utförs såväl på nationell som på lokalnivå. På lokalnivå använder man inom polisen det officiella verktyget LUPP, som utöver prognosmodellering även erbjuder möjligheter till analys och uppföljning. Syftet med denna uppsats är dels att redogöra för denna metod, och dels att göra en jämförelse med mer välkända statistiska modeller med avseende på prognoskvalitén. Jämförelsen bygger på antalet anmälda bilinbrott och görs för Skånes samtliga fem polisområden. Analysen baseras på månadsdata för perioden 1998-2003, där prognos och verkligt utfall kommer att analyseras för respektive modeller och polisområde. Analysens huvudsakliga mål är att komma till en slutsats huruvida polisen egen prognosmodell kan konkurrera med mer vetenskapliga metoder, och därigenom också identifiera en bästa metod. Det sekundära målet är att se om det finns några skillnader mellan de olika polisområdenas prognoskvalité, med avseende på användandet av LUPP-prognoser.

Efter utförda tester kan man konstatera att man inte med statistisk säkerhet kan konstatera någon bästa modell. Varken test av samtliga polisområden ackumulerat eller test av enskilt polisområde, visade på några signifikanta skillnader mellan prognosmetoderna. Med utgångspunkt från genomsnittligt prognosfel tillsammans med karaktären på de olika metoderna, kan emellertid metoden för Tidsserieuppdelning vara den prognosmetod som kan rekommenderas. Den avslutade analysen för den officiella prognosmodellen LUPP, visade att prognoser enligt denna metod visade sig fungera bäst för polisområdet norra Skåne (PONS).

## 2. Syfte

Syftet med denna uppsats är att göra en jämförelse mellan polisens prognosmodell, lokal uppföljnings & prognosprocedur (LUPP) - och de mer välkända metoderna:

- Box Jenkins ARIMA modeller
- Tidsserieuppdelning (Decomposition method)

Denna jämförelse har sin utgångspunkt i den prognoskvalité som varje modell uppvisar för perioden Jan - Dec 2004, utifrån brottskategorin bilinbrott. Resultatet från denna studie skall således visa vilken metod, respektive polisområde, som uppvisar den högsta prognoskvalitén.

### 3. Kodning av brott

#### 3.1 Generellt om kodning av brott

För samtliga brottstyper finns det en kod. Dessa koder sammanställs och publiceras av Brottsförebyggande rådet (BRÅ) och är officiell för den svenska polisen. Kodningen ligger till grund bland annat för Sveriges officiella statistik men även för andra statistiska mätningar. En noggrann och konsekvent kodning är en förutsättning för att redovisning och uppföljning av brott sker korrekt. Detta i sin tur leder till en högre effektivitet med avseende på planering, prioritering och utnyttjande av polisens samlade resurser. Huvudregeln vid kodning av brott är att varje gärning endast skall knytas till *en* brottskod.

#### 3.2 Definition bilinbrott

Bilinbrott definieras juridiskt som stöld ur, och från motordrivet fordon. Bilinbrott är en underkategori till benämningen bilbrott, där även brottstypen biltillgrepp (bilstöld) ingår. Med motorfordon inkluderas motordrivna fordon, släpfordon och husvagn. Motordrivna arbetsredskap inkluderas däremot inte i benämningen motorfordon. Bilinbrott kategoriseras med kod 0840.

### 4. Metod

#### 4.1 Introduktion till valda metoder

För att skapa de prognosvärden som skall ligga till grund för analysen av prognoskvalitén, kommer följande metoder att användas:

- Lokal uppföljnings & prognosprocedur (härefter benämnd LUPP)
- Box Jenkins ARIMA modeller
- Tidsserieuppdelning

Prognoserna skapas utifrån en tidsserie bestående av antalet bilinbrott under perioden 1998-2003. Prognoserna görs sedan månadsvis (Jan-dec.) för 2004, där sedan avvikelser mellan prognos och utfall för varje enskild månad räknas fram. Detta prognosfel (definieras vid 4.4) ligger sedan som grund för jämförelsen. Med avseende på prognosernas längd kan de beskrivas som 12-stegsprognoser, vilket därmed ger en 12-månaders horisont. Detta innebär att prognosen för exempelvis januari 2004, har skapats vid tidpunkten januari 2003. För att skapa prognoser för hela 2004 har följaktligen 12st 12-stegsprognoser formulerats. Valet att använda en 12-månaders prognoshorisont grundar sig i den organisatoriska verklighet som polisen lever i. Resurserna budgeteras och fördelas årsvis, vilket gör att uppskattningar om framtiden bör vara mer långsiktiga än kortsiktiga. Polisens organisation har således inte möjligheten att på kort sikt omfördela sina resurser i någon större omfattning, vilket exempelvis medför att prioriterade områden bör identifieras så tidigt som möjligt.

Valet att använda ARIMA modeller och tidsserieuppdelning i jämförelsen med LUPP, har sin utgångspunkt i metodernas uppbyggnad. ARIMA modellerna är de i sammanhanget mest teoretiska och står därmed i en klar kontrast till LUPP metoden. Metoden för tidsserieuppdelning faller någonstans mitt emellan, och har följaktligen en naturlig plats i denna studie. Modellvalet är också intressant i det avseendet att ARIMA modeller framför allt anses vara lämpliga för tidsserier innehållande parametrar som fluktuerar, medan modellen för tidsserieuppdelning främst är lämplig för stabila tidsserier. Det är således intressant att se i vilken omfattning modellernas karakteristik påverkar resultatet. Gemensamt för samtliga metoder är dock att de är univariata, det vill säga bygger sina prognoser enbart utifrån tidigare värden på tidsserien.

#### 4.2 Genomförande och redovisning

De analyser som har utförts i syfte att identifiera lämpliga prognosmodeller har genomförts med hjälp statistisk programvara. Detta arbete kan beskrivas som en stegvis process där vägen till den slutgiltiga modellen ibland är lång. För att undvika en alltför omfattande redogörelse av olika modeller, redovisas i bilagan därför endast karakteristik för de slutgiltigt valda modellerna. Med avseende på de tester som genomförs så provas de i samtliga fall på fem procentsnivån.

### 4.3 Allmänt om prognoser

Många beslut som fattas inom såväl näringslivet som den offentliga sektorn är starkt relaterade till framtiden. Gemensamt för dessa frågeställningar är att de beslut som fattas idag, får konsekvenser i det längre perspektivet. Gemensamt är också att samtliga beslut har ett framtidsperspektiv, och innehåller därför alltid en varierande grad av osäkerhet. För att minimera osäkerheten kring dessa beslut använder man sig därför av prognoser. Prognoser kan definieras som en uppskattning av framtida händelser, och kan skapas såväl utifrån ett kvalitativt som ett kvantitativt perspektiv. Man kan alltså, förutom att skapa prognoser utifrån traditionell sifferdata som exempelvis tidsserier, även formulera prognoser utifrån kvalitativ information som attityder, expertkunskap etc.

Prognosernas längd kan variera men en vanligt förekommande prognoshorisont är *ett* år. Detta kan främst förklaras av att den vanligaste budgetperioden är ett år, för vilken prognoserna skall försöka förutse och beskriva utvecklingen för det nästkommande budgetåret. Inom det offentliga rummet är också prognoser med längre perspektiv vanliga. Ett exempel på detta är de s.k. långtidsutredningarna (LU), som genomförs av statliga myndigheter. Dessa utredningar kan beskrivas som komplexa analysmodeller där ett flertal variabler beaktas utifrån olika antaganden. I dessa utredningar brukar tidsperspektivet var fem år och längre. Ett exempel på prognoser med kortare perspektiv är de olika konjunkturbarometrar som frekvent publiceras. Där ett exempel är Konjunkturinstitutets ”hushållens inköpsplaner”, en undersökning som genomförs månadsvis utifrån telefonintervjuer med ett urval av Sveriges befolkning.

Trots ett frekvent användande av prognoser är tilltron till detta verktyg högst varierande. Det grundar sig i framför allt två orsaker. Det första är att prognoser ibland slår fel, att kvalitén kort och gott upplevs som dålig. Det andra skälet har sin grund i användarperspektivet. Ofta presenteras prognoser alltför okritiskt och förbiser ofta den osäkerhet som alltid finns. Det finns också en risk att prognoserna blir heliga riktlinjer i den meningen att löpande information förkastas, vilket gör att den taktiska styrningen försvagas till förmån av den strategiska. Oavsett åsikt, så är trots allt prognoser det verktyg företag och andra organisationer har att tillgå för att förutspå framtiden. Följaktligen är prognoser något som alltid kommer att finnas, och som dessutom kommer att vara fortsatt omtvistade.



#### 4.4 Prognosfel

Prognosfelet kan definieras som differensen mellan det sanna värdet och prognosen.

Prognosfelet för en enskild prognos ( $\hat{y}$ ) är allmänt:  $e_t = (y_t - \hat{y}_t)$  där  $y$  är det sanna värdet.

Denna typ av prognosfel har dock den svagheten av negativa och positiva avvikelser jämnar ut varandra vid summering  $t_1, t_2, \dots, t_n$ . Av den anledningen kommer istället följande mått att ligga till grund för analyserna, prognosfel:  $\sqrt{e_t^2}$

Vid redovisning av prognosmodeller utifrån metoden för Tidsserieuppdelning, kommer även nedan mått att användas. Dessa mått beskriver hur väl den skattade modellen anpassar sig till observationerna. Samtliga mått skall tolkas som att ju lägre kvot, desto bättre anpassning.

$$\text{MAPE: } \frac{\sum |(y_t - \hat{y}_t) / y_t|}{n/100} \quad \text{Formel 4.4.1}$$

$$\text{MAD: } \frac{\sum |y_t - \hat{y}_t|}{n} \quad \text{Formel 4.4.2}$$

$$\text{MSD: } \frac{\sum (y_t - \hat{y}_t)^2}{n} \quad \text{Formel 4.4.3}$$

#### 4.5 Introduktion tidsserier

De statistiska metoder som denna uppsats behandlar utgår från teorin att utifrån ett kvantitativt perspektiv, skapa prognoser baserade på så kallade diskreta tidsserier. En tidsserie kan definieras som en kronologisk följd av observationer för en specifik variabel. Tidsserierna kan även delas upp till att vara antingen diskreta eller kontinuerliga. En diskret tidsserie mäts utifrån ekvidistanta tidpunkter (samma avstånd mellan punkterna)  $t_1, t_2, t_n$  d.v.s. vi har observationerna  $y_1, y_2, \dots, y_n \rightarrow y_t, t = 1, 2, \dots, n$ . Exempel på sådana observationer kan vara månatlig försäljning av en vara medan en kontinuerlig tidsserie bygger på kontinuerlig data som exempelvis temperaturer. En kvantitativ prognosmodellering beskriver metoder som bygger på historisk, numerisk data. Kvantitativa metoder kan delas in i två huvudgrupper, univariata respektive kasuala modeller. Där en univariat modell skapar prognoser för  $y_t$  endast utifrån tidigare värden  $y; y_{t-1}, y_{t-2}, \dots, y_{t-n}$ . En kasual modell (ekonometrisk modell) innebär att man skapar prognoser för  $y_t$  med hjälp av en eller flera hjälpvariabler  $x_t, z_t$ .

#### 4.6 Tidsseriens innehåll

Det finns ingen universal beskrivning av tidsseriens innehåll. En traditionell indelning och därigenom också en vanlig definition av tidsseriens olika delar, kan emellertid vara följande.

Trend:

En trend beskriver tidsseriens långsiktiga utvecklingsmönster som antingen kan vara långsiktigt växande eller långsiktigt avtagande. Vid en trend bortser man ifrån kortsiktiga och tillfälliga förändringar. Vid tidsserieanalys är trenden intressant ut två aspekter. För det första så innebär en skattad trend en möjlighet till att formulera prognoser. För det andra så kan man genom att eliminera trenden, bidra till en enklare identifikation och tolkning av tidsseriens övriga komponenter som exempelvis säsongvariation.

Cyklisk variation:

Den cykliska variationen innebär rörelser kring trendnivåerna och kan beskrivas som ett regelbundet återkommande mönster. I jämförelse med säsongvariation, är dock inte detta mönster så regelbundet och så kortvarigt. De cykliska variationerna kan uppvisa ett mycket varierande mönster i avseende på längd och omfattning. För exempelvis ekonomiska serier, kan man identifiera cykler om längder 3-5 år. Allmänt kan man dock säga att cyklisk variation kan ha längder mellan 2-10 år. Ett mer vedertaget begrepp för cyklisk variation är konjunktur.

Säsongvariation:

Denna typ av variation bygger på periodiska skiftningar inom ett kalender år. Dessa upprepas årligen enligt ett regelbundet återkommande mönster. Säsongvariation har vanligtvis sitt ursprung i väder och vanor, seder och bruk. Säsongvariation har två särarter, additiv och multiplikativ. Med additiv menas att säsongvariationen är konstant. Medan multiplikativ innebär att säsongvariationen är ökande respektive minskande, det vill säga är mer relaterad till övriga värden under tidsserien.

Slumpvariation:

Dessa variationer bygger på händelser som inte kan förklaras eller knytas till något regelbundet mönster. Dessa kan bland annat ha sin grund i slumpartade händelser, som exempelvis jordbävningar, olyckor, krig och strejker.

#### 4.7 Att beakta vid skapande av prognoser

Tidsseriens olika komponenter kan förekomma i ett flertal olika kombinationer, av den anledningen är det därför omöjligt att hitta *en* bästa prognos modell. Den största utmaningen är således att hitta den modell som är bäst anpassad till den aktuella tidsserien och det mönster serien uppvisar. Först efter att en modell med god anpassning har identifierats, kan modellens parametrar skattas och prognoser formuleras.

#### 4.8 Tidsserieuppdelning

##### 4.8.1 Inledning

Tidsserieuppdelning används för att prognostisera tidsserier med trend och säsongskomponenter. Metoden anses ge bra modeller på tidsserier där ingående parametrar är stabila över tiden. Principen med denna metod är dela upp tidsserien i flera delkomponenter – trend, säsong, cykliskvariation och slump. Dessa beståndsdelar skattas därefter och beskriver därmed tidsserien, vartefter skattningarna, om de överliggande parametrarna hålls konstanta, kan användas för att göra prognoser. Metoden innebär följaktligen att modellen kan beskrivas utifrån dess ingående parametrar:

$Y_t$  = Observerat värde i tidsserien vid tid  $t$

$T$  = Trendkomponent

$S$  = Säsongskomponent

$C$  = Cyklisk komponent

$R$  = Slump komponent

Man kan vidare dela upp denna metod i:

- Multiplikativ modell
- Additiv modell

Den multiplikativa modellen används för tidsserier med ökande alternativt minskande säsongvariation och kan formuleras som:  $Y_t = T_t * S_t * C_t * R_t$

Additiv modell är bäst lämpad för tidsserier som uppvisar konstant säsongvariation. Den additiva modellen beskrivs som:  $Y_t = T_t + S_t + C_t + R_t$

Modellskattning enligt Tidsserieuppdelning kan göras i olika programpaket, bland annat i Minitab. I Minitab kan man utöver valet av multiplikativ eller additiv modell, även anpassa modeller utan trendkomponent. För att avgöra huruvida tidsserien innehåller en trend, kan man skapa prognoser utifrån både en multiplikativ och en additiv modell. Är prognoserna likartade, är detta ett tecken på att trenden är svag. Valet mellan additiv eller multiplikativ modell, kan underlättas genom att studera diagram för tidsserien (plot). De flesta tidsserier med säsongvariation antas emellertid uppvisa ett multiplikativt mönster. För att dela upp seriens olika komponenter börjar man med att skatta trendkomponenten. För att åstadkomma detta kan man använda antingen metoden för regressionsanalys eller metoden för glidande medelvärde. I denna rapport har, ty enklare prognoshantering, regression använts för att skatta tidsseriens komponenter. När trenden har skattats kan sedan resterande komponenter identifieras.

#### 4.8.2 Skattning av trendkomponent

För att skatta trenden anpassas tidsserien till en rät linje enligt minsta-kvadrat-metoden. Denna metod går ut på att identifiera den linje som minimerar kvadratsumman:

$$\sum (y_i - \hat{y}_i)^2 \quad \text{Formel 4.8.1}$$

Den skattade ekvationen används sedan också för att göra prognoser. För att skatta trenden används nedan formler. Låt tidsserien vara den beroende variabeln ( $y$ ) och observationernas löpnummer ( $t_1, t_2, \dots, t_n$ ), vara förklarande variabler ( $x$ ).

$$b = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum x_i^2 - (\sum x_i)^2} \quad \text{Formel 4.8.2}$$

$$a = \frac{\sum y_i}{n} - b * \frac{\sum x_i}{n} = \bar{y} - b\bar{x} \quad \text{Formel 4.8.3}$$

För mer information om lämpliga tillvägagångssätt för att skatta tidsseriens övriga komponenter, hänvisas till Flemming m.fl. (2000).

## 4.9 Box-Jenkins

Box-Jenkins metoden innebär ett mer systematiskt arbetssätt för att skapa prognoser och ger möjlighet till att skapa prognoser utifrån alla typer av tidsserier. Metoden anses dock vara mest lämplig för tidsserier med relativt snabba förändringar. Box-Jenkins är en univariat metod ty prognoser baseras endast på tidigare värden av  $y_t$ .

Arbetet med att forma en slutgiltig modell kan delas in i fyra steg:

1. Preliminär identifiering: Tidsseriedata används för att bestämma en preliminär Box-Jenkins modell.
2. Estimat: Tidsseriedata används för att skatta parametrar i den preliminära modellen.
3. Diagnostik: Olika metoder används för att kontrollera den preliminära modellen och dess parametrar i syfte att säkerställa att den är adekvat och riktig. Under detta steg kan en ny preliminär modell antas.
4. Prognostisering: När den slutgiltiga modellen är vald, används den för att prognostisera framtida tidsserie data.

När man använder Box-Jenkins metodologin utgår man från kraven om stationära och invertibla tidsserier. En invertibel Box-Jenkins modell antyder att vikter avtar med tiden. Nutida observationer har således en större inverkan på prognosen än äldre observationer. En praktisk beskrivning av detta är till exempel att parametrarna i en MA-modell, tillsammans inte skall summera till värde över ett. En stationär tidsserie kan definieras som en tidsserie där viktiga egenskaper såsom väntevärde och varians, är konstanta över tiden. En tidsserie som inte kan beskrivas som stationär kan oftast, med olika metoder, transformeras till stationäritet.

### 4.9.1. Olika Box-Jenkins modeller

Box-Jenkins familjen innehåller flera olika typer av modeller. Två av dessa är de autoregressiva (AR) och så kallade glidande medelvärde (MA) modellerna.

#### AR modell av ordningen P

För denna modell gäller:

$$Z_t = \delta + \Phi_1 Z_{t-1} + \Phi_2 Z_{t-2} + \dots + \Phi_p Z_{t-p} + a_t \quad \text{där } p = 1, 2, \dots, n.$$

Där  $Z_t$  är en funktion av tidigare värden på  $z_{t-1}, z_{t-2}, \dots, z_{t-p}$  samt en slumpterm,  $a_t$ . Parametern ( $\Phi$ ) är okänd och skattas utifrån  $z_t, z_{t-1}, z_{t-2}, \dots, z_{t-p}$  där  $p = 1, 2, \dots, n$ .

Konstanten ( $\delta$ ) är en funktion av väntevärdet där:

$$\delta = \mu(1 - \Phi_1 - \Phi_2 - \dots - \Phi_p) \quad p=1, 2, \dots, n$$

$a_t$  är en normalfördelad slumpvariabel med varians och väntevärde = 0,  $a_1, a_2, \dots, a_n$ .

Dessa variabler antas också vara oberoende

### MA modell av ordningen q

För denna modell gäller:

$$Z_t = \delta + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q} \quad \text{där } q = 1, 2, \dots, n.$$

Här använder man utöver  $a_t$  (som används i samtliga Box-Jenkins modeller) även tidigare värden på  $a_t; a_{t-1}, a_{t-2}, \dots, a_{t-q}$ . Parametern ( $\theta$ ) är okänd och skattas utifrån  $z_t$  från  $a_{t-1}, a_{t-2}, \dots, a_{t-q}$ .

$a_t$  är en normalfördelad slumpvariabel med noll varians och väntevärde 0 där samtliga  $a_t$  ( $a_1, a_2, \dots, a_n$ ) antas vara oberoende av varandra.

Konstanten ( $\delta$ ) är en funktion av väntevärdet där:

$$\delta = \mu(1 - \Phi_1 - \Phi_2 - \dots - \Phi_p) \quad p=1, 2, \dots, n$$

### 4.9.2 Identifikation av lämplig modell

För att uppnå kravet på stationaritet finns olika metoder för transformering. För att få en uppfattning om tidsseriens (observationernas) fördelning, kan man genomföra en Box-Cox transformation. Denna metod är lämplig för att identifiera skeva dataserier med exempelvis stor varians. För att stabilisera variansen och därmed också förbättra tidsseriens stationaritet, ger Box-Cox transformationen riktlinjer för lämplig s.k. för-transformation. Vid en Box-Cox transformation erhålls ett lambdavärde som är knutet till en rekommenderad transformering. Där exempelvis ett lambdavärde nära 0, innebär att tidsseriens bör logaritmeras. En annan metod för att skapa stationära serier är transformera tidsserien genom att ta så kallade differenser.

Första differensen av tidsseriedata  $y_1, y_2, \dots, y_n$  är

$$Z_t = y_t - y_{t-1} \text{ där } t = 2, \dots, n$$

Andra differens av tidsseriedata  $y_1, y_2, \dots, y_n$  är  $Z_t = (y_t - y_{t-1}) - (y_{t-1} - y_{t-2})$

$$= y_t - 2y_{t-1} + y_{t-2} \text{ där } t = 3, 4, \dots, n$$

För att sedan kontrollera om en tidsserie kan antas vara stationär, kan man betrakta autokorrelationen (ACF) och den partiella autokorrelationsfunktionen (PACF).

Autokorrelation:

Formel 4.9.1

$$r_k = \frac{\sum_{t=b}^{n-k} (z_t - \bar{z})(z_{t+k} - \bar{z})}{\sum_{t=b}^n (z_t - \bar{z})^2}$$

där

Formel 4.9.2

$$\bar{z} = \frac{\sum_{t=b}^n z_t}{n - b + 1}$$

Formel 4.9.3

Standardfelet för  $r_k$

$$s_{rk} = \frac{\left(1 + 2 \sum_{j=1}^{k-1} r_j^2\right)^{1/2}}{(n - b + 1)^{1/2}}$$

$t_{rk}$  – kvot enligt

Formel 4.9.4

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

Detta mått mäter den linjära relationen mellan tidsserieobservationer uppdelade i lag  $k$  tidsenheter, det vill säga är tidsförskjutna  $t_1, t_2, \dots, t_n$ .  $R_k$  är ett värde som alltid är mellan  $-1$  och  $1$ . Ett värde nära  $1$  indikerar att observationerna har ett positivt linjärt samband, medan ett värde på  $-1$ , tyder på ett negativt linjärt samband. Vid bedömning om eventuell stationaritet försöker man identifiera  $t$ -kvoter som vanligtvis är större än  $1,6$  när det gäller få lags ( $k=1,2,3$ ). När man bedömer stationaritet vid flera lags ( $K>3$ ), söker man ett  $t$ -kvot värde som är större än  $2$ . För att sedan slutgiltigt bedöma om en tidsserie är stationär, måste samtliga lags

med respektive t-värden analyseras. Där syftet är att undersöka om det föreligger någon så kallad "cut-off" det vill säga när autokorrelationsfunktionen efter lag med hög t-kvot, avtar snabbt alternativt relativt snabbt, mot noll.

Partiell autokorrelationsfunktion:

Detta mått används också för att identifiera lämplig Box-jenkins modell. Skillnaden mellan ACF och PACF är att i en partiellautokorrelationsfunktion eliminerar man effekterna av tidsseriens emellanliggande observationerna bestående av lag 1, 2, ..., n.

Den partiella autokorrelationen för lag  $k$  är =

$$r_{kk} = \begin{cases} r_1 & \text{if } k = 1 \\ \frac{r_k - \sum_{j=1}^{k-1} r_{k-1,j} r_{k-j}}{1 - \sum_{j=1}^{k-1} r_{k-1,j} r_j} & \text{if } k = 2, 3, \dots \end{cases} \quad \text{Formel 4.9.5}$$

Där

$$r_{kj} = r_{k-1,j} - r_{kk} r_{k-1,k-j} \quad \text{för } j = 1, 2, \dots, k-1$$

2.. Std.felet av  $r_{kk}$  är

$$s_{r_{kk}} = 1/(n - b + 1)^{1/2} \quad \text{Formel 4.9.6}$$

3.  $t$ -kvoten är

$$t_{r_{kk}} = \frac{r_{kk}}{s_{r_{kk}}} \quad \text{Formel 4.9.7}$$

#### 4.9.3 Anpassning av Box-Jenkins modeller

Möjligheten att utifrån en given korrelationsfunktion (ACF, PACF) identifiera en lämplig ARIMA modell, har sitt ursprung i det förutbestämda, teoretiska korrelationsmönster som varje enskild ARIMA modell har. Dessa teoretiska mönster - teoretisk autokorrelation (TAC) och teoretisk partiell autokorrelation (TPAC), ger riktlinjer om vilken ARIMA modell som är lämplig för varje identifierad korrelationsfunktion.



## Exempel 4.9.3

För en MA modell av ordningen 2 (MA(2)) gäller :  $Z_t = \delta - \theta_1 a_{t-1} - \theta_2 a_{t-2} + a_t$

där den givna, teoretiska korrelationsfunktionen följande utseende:

TAC: Har korrelation  $> 0$  vid lag 1 samt lag 2, och har korrelation 0 för påföljande lag

TPAC: Korrelationsfunktionen är exponentiellt avtagande

För att bekräfta en MA(2) modell måste den analyserade tidsserien uppvisa följande korrelationsmönster:

ACF: Visar på ”cut-off” efter lag 2 och avtar sedan snabbt

PACF: Uppvisar ingen ”cut-off” utan visar ett avtagande mönster

När man identifierar ARIMA modeller, gör man det således genom att analysera både ACF och PACF. I diagrammet 4.9.3a nedan, ser man att autokorrelationsfunktionen har en tydlig ”cut-off” efter lag 2 och därmed antas tidsserien vara stationär. Korrelationsmönstret i den partiella autokorrelationsfunktionen (diagram 4.9.3b), kan beskrivas som avtagande.

Diagram 4.9.3a Autokorrelation

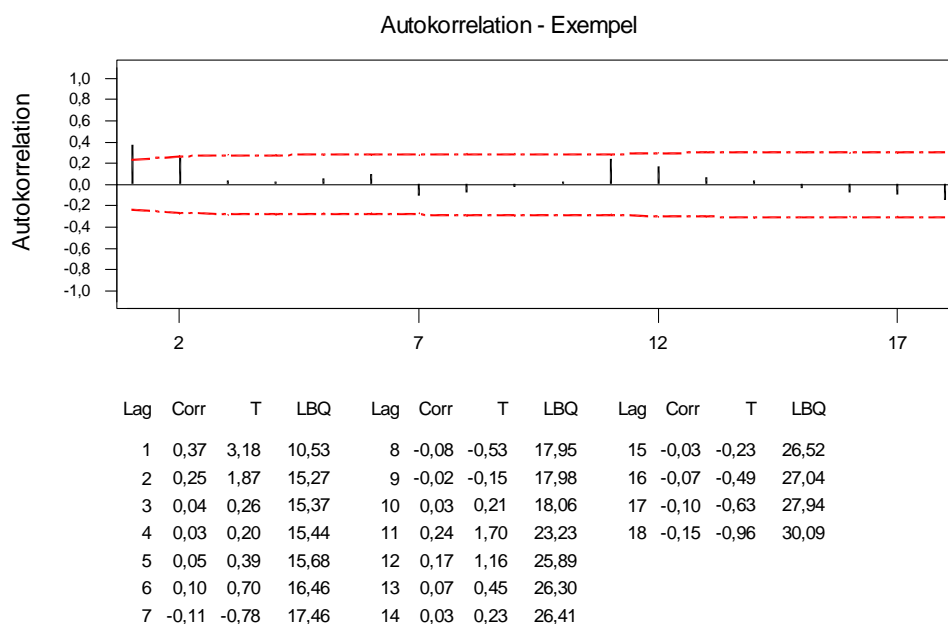
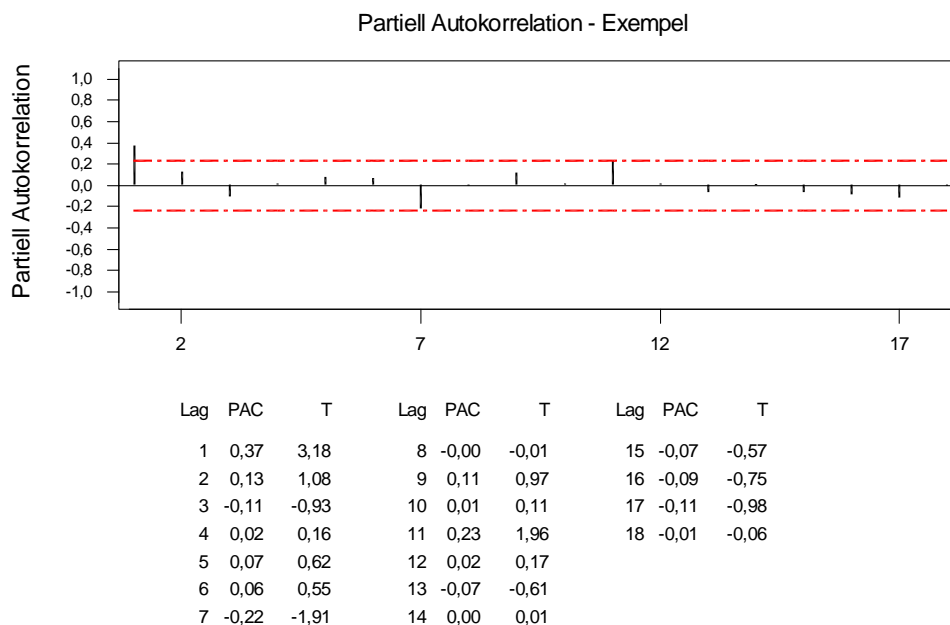


Diagram 4.9.3b Partiell autokorrelation



Utifrån denna analys där ACF visar på en tydlig ”cut-off” efter lag 2 samtidigt som PACF visar på ett avtagande mönster, kan man identifiera en MA-modell av ordning 2, MA(2). Sammanfattningsvis kan man konstatera att en ”cut-off” i autokorrelationsfunktionen tillsammans med långsamt avtagande lags i den partiella autokorrelationsfunktionen, ger MA-modeller. Omvänt kan det även bevisas att AR-modeller ges av ”cut-off” i PACF i kombination med långsamt avtagande Lags i ACF. Ett korrelationsmönster som visar på tydlig ”cut-off” i såväl ACF som PACF, kan peka på att en kombinerad modell bestående av såväl MA som AR komponenter kan vara lämplig.

#### 4.9.4 Diagnostik

När en ARIMA-modell har anpassats och skattats, bör den kontrolleras för att slutligt kunna fastställas. För exemplet 4.9.3 ser vi att samtliga parametrar inklusive konstanten är signifikanta med höga t-värden. Vi ser även att anpassningen i form av ett Ljung-box test är god ty testet ger höga p-värden. Detta test har sin utgångspunkt i residualernas fördelning och indikerar i detta exempel på ett oberoende. Oberoende residualer är en förutsättning för att modellen skall vara adekvat och tillämplig för statistisk inferens.

Diagnostik för exempel 4.9.3:

Type	Coef	SE Coef	T	P
MA 1	-0,3203	0,1149	-2,79	0,007
MA 2	-0,3060	0,1151	-2,66	0,010
Constant	199,909	5,256	38,04	0,000
Mean	199,909	5,256		

Number of observations: 72

Residuals: SS = 51977,7 (backforecasts excluded)

MS = 753,3 DF = 69

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

Lag	12	24	36	48
Chi-Square	11,1	15,3	28,2	45,3
DF	9	21	33	45
P-Value	0,271	0,808	0,703	0,459

I detta exempel erhålls därmed (avrundat till två decimaler) följande modell:

$$Z_t = 199,91 - (-)0,32a_{t-1} - (-)0,31a_{t-2} + a_t$$

Skattningarna i ARIMA baseras på maxi-metoden, vilken har asymptotiska egenskaper. Begreppet asymptotiskt har sin grund i centrala gränsvärdessatsen och innebär praktiskt att när  $n$  går mot oändligheten, så antas till exempel det skattade värdet vara det sanna värdet. Detta faktum beskrivs allmänt som att skattningen är asymptotiskt väntevärdesriktig. En skattning kan också vara asymptotiskt effektiv, detta innebär att variansen går mot noll när  $n$  går mot oändligheten.

För en maxiskattning av  $\theta$  gäller alltså:

1. Skattningen är konsistent, d.v.s.  $\theta$  konvergerar mot en en-punktsfördelning, då  $n$  går mot oändlighet.
2. Skattningen är asymptotiskt väntevärdesriktig

3. Skattningen är asymptotiskt effektiv så minsta möjliga varians erhålls då  $n$  går mot oändligheten.
4. Skattningen är asymptotiskt normalfördelad då  $n$  går mot oändligheten.

Man bör även grafiskt undersöka residualerna för att bekräfta en vald modell. I denna analys skall inga värden överträda gränslinjerna som representeras av  $2 * \text{MSE}$ . Eventuella överträdelser innebär att modellen har bristfällig anpassning och således bör justeras. Utifrån nedan grafer (diagrammen 4.9.4a-b) kan man konstatera att ingenting tyder på att den valda modellen skulle vara olämplig, ty gränslinjerna är obrutna för samtliga lags.

Diagram 4.9.4a Autokorrelation - residualer

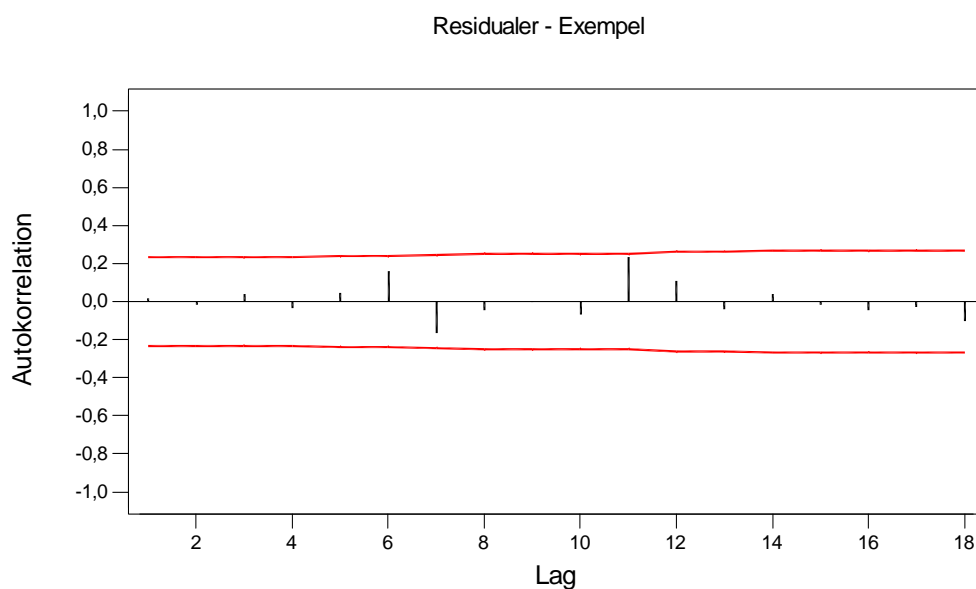
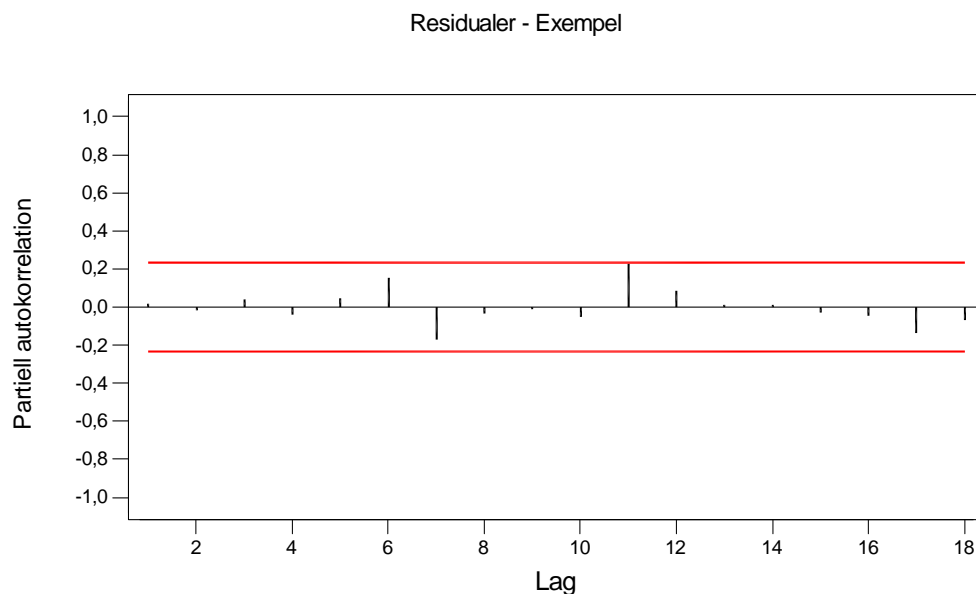


Diagram 4.9.4b Partiell autokorrelationsfunktion – residualer



#### 4.9.5 Box-Jenkins säsongmodeller

För tidsseriedata som innehåller säsongvariation, kan man anpassa Box-Jenkins modeller som innehåller en eller flera säsongparametrar. Säsongvariationen kan anta olika längder, vanligast förekommande är säsongslängd 6 och 12 för månadsdata och säsongslängd 4 för kvartalsdata. Säsongvariationen kan också vara konstant (additiv) eller stigande, respektive avtagande (multiplikativ). En så kallad SMA modell identifieras genom en tydlig ”cut-off” på säsongslags (Oftast Lag 4/6/12) i autokorrelationsfunktionen, samt ett mönster i den partiella autokorrelationsfunktionen som visar på att lags som följer efter säsongslag, långsamt avtar. Vid autoregressiva säsongmodeller (SAR) modeller gäller det motsatta. Det kan även förekomma blandande modeller där även parametrar utan säsong skall ingå. Dessa kan identifieras genom att korrelationsmönstret har en tydlig ”cut-off” även på icke-säsongslag. En cut-off i ACF rekommenderar en MA-komponent, medan en ”cut-off” i PACF, visar på att en AR-komponent bör ingå i modellen.

Ett grundkrav för att skapa säsongmodeller är analogt med tidsserier utan säsongvariation, kravet på stationaritet. Dessutom bör säsongvariationen vara konstant över tiden för att skapa en modell med god anpassning. Säsongmodeller baseras på AR(p) respektive MA(q) modeller, och definieras som

SAR(p)

Där L = säsongslängd

$$Z_t = \dots \delta + \Phi_{1,L} Z_{t-L} + \Phi_{2,L} Z_{t-2L} + \dots + \Phi_p Z_{t-p} + \dots + \Phi_{p,L} Z_{t-pL} + a_t$$

SMA(q)

$$Z_t = \dots \delta + a_t - \theta_{1,L} a_{t-L} - \theta_{2,L} a_{t-2L} - \dots - \theta_{Q,L} a_{t-QL}$$

För kombinerade MA och SMA respektive AR och SAR modeller, skall även multiplikativa termer adderas till den skattade modellen. Multiplikativa termer bildas utifrån modellens parametrar och vid exempelvis en kombinerad MA(1) och SMA(1) modell med säsongslängd 12, skapas den multiplikativa termen enligt följande:

Modellens parametrar är:  $-\theta_1 a_{t-1}$  och  $-\theta_{1,12} a_{t-12}$  vilket ger  $(-\theta_1)(-\theta_{1,12}) a_{t-13}$  som ger den multiplikativa termen:  $\theta_1 \theta_{1,12} a_{t-13}$

$Z_t = y_t - y_{t-12}$  adderas sedan med den multiplikativa termen vilket ger:  $Z_t = y_t - y_{t-12} + y_{t-13}$

Där prognoser för  $y_t$  kan skapas enligt modellen:

$$y_t = y_{t-12} - y_{t-13} - \theta_1 a_{t-1} - \theta_{1,12} a_{t-12} + \theta_1 \theta_{1,12} a_{t-13} + a_t$$

Diagnostik genomförs analogt med tidigare beskrivna metoder för AR och MA modeller utan säsongskomponenter.

#### 4.9.6 Prognosmodellering utifrån Box-Jenkins modeller

För att skapa prognoser utgår man från den modell som man har identifierat och genom diagnostik, även verifierat. Vid användandet av statistisk programvara ges prognoser med automatik utifrån valfria startpunkter och horisonter (valfritt antal steg). En manuellt framtagen en-stegsprognos, det vill säga en prognos för tidpunkten  $Y_{t+1}$ , kan utifrån exemplet 4.9.3 skapas enligt följande.

MA(2) modellen definierades som:  $Z_t = \delta + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2}$

och skattades till:  $Y_t = 199,91 + a_t - (-)0,32a_{t-1} - (-)0,31a_{t-2}$

Prognosen för tidpunkten  $Y_{t+1}$  blir således:  $199,91 + 0 - (-)0,32 * -4,5 - (-)0,31 * -3,3 = 197,447$   
 Givet att residual  $a_{t-1} = -4,5$ , residual  $a_{t-2} = -3,3$ . Residualen  $a_t$  blir 0 då residualen vid denna tidpunkt inte kan beräknas (verkligt värde för  $Y_{t+1}$  saknas.) Man kan även skapa ett konfidensintervall för det prognostiserade värdet. Ett 95-procentigt konfidensintervall för prognosen  $y_{t+1}$  skapas enligt:

$$\hat{y}_t = \pm 1,96 * \sqrt{MS} \text{ vilket ger intervallet } 197,447 \pm 27,44$$

(MS (här:753,3) ges per automatik vid skattningen av modellens parametrar, se diagnostik vid punkt 4.9.4)

## 4.10 LUPP

### 4.10.1 Beskrivning av LUPP

Metoden ”lokal uppföljnings & prognosprocedur i problemorienterat arbete mot brott” – LUPP, är en metod utvecklad av Brottsförebyggande rådet (BRÅ). I polisarbete som definieras som problemorienterat, behöver polisen ha tillgång till bra metoder och verktyg för att kunna kartlägga och analysera brottsligheten. Syftet med att metodiskt kartlägga och analysera den anmälda brottsligheten är att hitta ett mönster, mönster som också kan ge en förutsägelse - prognos, av vilken brottslighet som förväntas under den närmaste framtiden. Med LUPP skapar man prognoser och med prognosernas hjälp ska man sedan kunna planera arbetet. LUPP erbjuder dessutom möjligheten att kunna följa upp de åtgärder som vidtas för att motverka olika kategorier av vardagsbrott. Med vardagsbrott menas ständigt återkommande brottslighet, primärt olika typer av våld och tillgreppsbrott. LUPP är inte en statisk modell utan skall utöver att erbjuda grundläggande förutsättningar för analys och prognosarbete, även ge utrymme för ett stort mått av interaktivitet. LUPP används också med fördel på lokalnivå och är ej tänkt att användas som ett nationellt prognosverktyg.

### 4.10.2 LUPP i teorin

Genom att använda LUPP skapar man på ett relativt enkelt sätt prognoser genom att observera antal anmälda brott bakåt i tiden. En central utgångspunkt i utvecklingen av LUPP är man utgår från det faktum att brott inte sker slumpmässigt. Speciellt när det gäller vardagsbrottsligheten finns det ett mönster om, var, och när, brotten inträffar. Varför

brottssituationer uppstår kan beskrivas bland annat genom rutinaktivitetsteorin. Denna teori utgår från det faktum att brottsligheten ofta styrs av människors vardagliga och rutinmässiga aktiviteter. Enligt detta sätt att se på brottsligheten, är det alltså det vardagliga livet som skapar förutsättningar för brottsligheten, vilket bland annat medför att sannolikheten för olika brott är högre för vissa tidpunkter och platser. Brottsligheten är också separerad utifrån geografien, vissa brottstyper är mer frekventa inom städers centrala delar kontra landsbygd och omvänt. Det finns även skillnader i en större skala mellan olika kommuner och län.

#### 4.10.3 Prognosframställning via LUPP

För att skapa prognoser för vald brottskategori används antalet anmälda brott tre år tillbaka i tiden. För att skapa en prognos för januari 2005, undersöker man antalet anmälningar för en specifik brottstyp för januari 2002, 2003 och 2004. Man erhåller alltså tre observationer där man sedan väljer medianvärdet. Detta värde blir prognosen för januari 2005. Man undviker därmed att använda medelvärden vilka är känsliga för eventuella extremvärden (onormalt höga eller låga observationer), vilka man hävdar är enstaka företeelser som inte återkommer. Att man dessutom jämför månad för månad, gör att man även beaktar eventuell säsongsvariation. Detta beaktar det faktum att en viss typ av brottslighet, exempelvis cykelstölder, är mer vanligt förekommande under sommarmånaderna.

Runt varje månadsprognos bildas sedan ett osäkerhetsintervall. Man använder sig av t-fördelningen (tvåsidigt,  $\alpha = 0,05$ ) och bildar således ett 90-procentigt konfidensintervall med följande formel:  $1,64 * \sqrt{m}$  (m = månadsprognosen)

Om månadsprognosen är 225 blir således intervallet +/- 24,6, vilket gör att prognosintervallet blir 200,4 – 249,6. Månadsprognoserna kan sedan adderas så att en årsprognos kan skapas. Runt detta värde kan också, analogt med ovan, ett intervall skapas.



## 4.11 Introduktion till statistiska tester

### 4.11.1 Parametriska metoder

Dessa metoder bygger sin validitet på att man känner fördelningen för de observationer som skall analyseras. De flesta parametriska metoder är baserade på normalfördelningen, men det finns även parametriska procedurer för andra fördelningar som exempelvis exponential- och Weibull fördelningen. Testmetoder med antagande om normalfördelning, innebär exempelvis att man kan pröva hypoteser och skapa konfidensintervall för stickprovets medelvärde och standardavvikelse. Då fördelningen är okänd men stickprovet stort, kan man dock utifrån centrala gränsvärdessatsen approximera observationerna som normalfördelade. Samtliga metoder som är approximativt tillämpliga även om något antagande (exempelvis antagandet om normalfördelning) inte är uppfyllt, betecknas som *robusta* mot respektive antagande. Detta innebär att parametriska metoder även är robusta (tillämpliga) då exakt normalfördelning inte föreligger. Detta innebär emellertid inte att en robust metod nödvändigtvis har hög styrka (tillförlitlighet), när stickprovet inte är normalfördelat. Då fördelningen är okänd, kan istället mer tillförlitliga tester genomföras med hjälp av icke-parametrisk statistik.

### 4.11.2 Icke-parametriska metoder

Vid icke-parametriska tester finns inget antagande om någon specifik fördelning. Detta innebär att de är robusta mot samtliga fördelningsantaganden och kan således tillämpas på alla stickprov, oavsett fördelning. Icke-parametriska tester är tillämpliga för såväl kvantitativ som för kvalitativ data. Några vanliga icke-parametriska tester baseras bl.a. på sannolikhetstabeller och tester baserade på tecken (negativa eller positiva avvikelse mellan parvisa observationer) eller ordningstal (rank). Vid situationer när även parametriska metoder är robusta, bör det test väljas som ger högst styrka. Vid exempelvis data liknande normalfördelning men med lättare svansar, är det parametriska testet lika, eller mer kraftfullt än de icke-parametriska testerna. Ett praktiskt exempel på sådana data är likformigt fördelat material som exempelvis undersökningsdata, där respondenten får besvara frågor utifrån en skala från 1 till 5 eller 1 till 7 etc. Vid ett datamaterial betecknat av tunga svansar som exempelvis Chi-två-fördelade variabler, har emellertid icke-parametriska metoder högre styrka. Ett exempel på ett sådant datamaterial är vid situationer då det förekommer extremvärden

## 4.12 Formulering av hypoteser

### 4.12.1 Undersökning av prognoskvaliteten - samtliga polisområden

För att undersöka prognoskvaliteten för samtliga polisområden ackumulerat, utförs en ensidig variansanalys. Syftet med detta test är att se om någon av prognosmetoderna ger signifikant bättre prognoser än de övriga. Den undersökta faktorn i detta test blir således variabeln ”prognosmetod”, som i sin tur tilldelas tre nivåer: ARIMA, LUPP och Tidsserieuppdelning. Varje nivå har 60 replikat, vilka representerar de 60st månatliga prognosfel som räknats fram för perioden januari till och med december 2004 för Skånes samtliga fem polisområden. Blir den undersökta faktorn signifikant (lågt P-värde), innebär detta att nollhypotesen kan förkastas. Vid förkastad nollhypotes kan man dra slutsatsen att valet av prognosmetod har signifikant betydelse för prognoskvaliteten för Skånes polisområden. För detta test ställs följande hypoteser:

$H_0$ : Prognoskvaliteten är lika för samtliga prognosmetoder

$H_1$ : Prognoskvaliteten skiljer sig åt för någon av metoderna

### 4.12.2 Undersökning av prognoskvaliteten - enskilt polisområde

För det enskilda polisområdet prövas prognoskvaliteten med det icke-parametriska testet Kruskal-Wallis. Detta test är lämpligt för att pröva hypoteser om lika populationer, vid tillfällen då man vill analysera fler än två k-grupper. Vid detta test delas prognosmetoderna in i tre grupper där sedan samtliga observationer (prognosfelen) tilldelas ett ordningstal (rang). Den lägsta observationen får rang nummer 1, den näst lägsta, rang nummer 2 o.s.v. Testet baseras således på observationernas rang och inte på själva värdet (prognosfelet). Det är dock enkelt att tolka relationen mellan prognosfel och rangtal. Ett lågt rangtal indikerar på ett relativt sätt litet prognosfel, och omvänt. Det summerade rangtalet för respektive grupp, ger således vägledning om prognosfelens relativa storlek. Kruskal-Wallis test är användbart vid små stickprov samt vid situationer där observationernas fördelning är okänd. Dessutom är det ett lämpligt test då man inte kan göra någon bedömning - se något värde, av en enskild observation. Vid användandet av Kruskal-Wallis test i denna rapport, görs vid några av testerna justeringar för så kallade ”ties”. Med begreppet ”ties” menar man observationer med

lika värde och som följaktligen tilldelas samma rangtal. Då antalet ”ties” är litet, är justering ej nödvändig. För detta test definieras följande hypoteser:

$H_0$ : Prognoskvaliteten är lika för samtliga prognosmetoder

$H_1$ : Prognoskvaliteten skiljer sig åt för någon av metoderna

Testvärdet ges av formel 4.12.1:

$$T = \frac{12}{N(N+1)} \sum_{i=1}^K \frac{R_i^2}{n_i} - 3(N+1) \quad \text{Formel: 4.12.1}$$

Där nollhypotesen förkastas om  $T > \chi^2_{0,05 (k-1)}$

Då antalet ”ties” är flera används istället formel 4.12.2:

$$T = \frac{1}{S^2} \left( \sum_{i=1}^k \frac{R_i^2}{n_i} - \frac{N(N+1)^2}{4} \right) \quad \text{Formel 4.12.2}$$

Där  $s^2$  skattas enligt

$$S^2 = \frac{1}{N-1} \left( \sum_{\text{rangtal}} R(X_{ij})^2 - N \frac{(N+1)^2}{4} \right) \quad \text{Formel 4.12.3}$$

Förkastas nollhypotesen kan grupperna sedan provas parvis, för att exempelvis identifiera den metod som har lägst rang och följaktligen också de bästa prognoserna. Dessa parvisa tester utförs med formel 4.12.4:

$$\left| \frac{R_i}{n_i} - \frac{R_j}{n_j} \right| > t_{1-(\alpha/2)} \left( S^2 \frac{N-1-T}{N-k} \right)^{1/2} \left( \frac{1}{n_i} + \frac{1}{n_j} \right)^{1/2} \quad \text{Formel 4.12.4}$$

Där  $S^2$  ges av:  $N(N+1)/12$

Formel 4.12.5

Konstanten T ges av Kruskal-Wallis testformel ovan. Signifikansen för testet ges av valfri nivå från T-fördelningen, N-k frihetsgrader.

#### 4.12.3 Undersökning av prognoskvalitén med avseende på LUPP

För att statistiskt utvärdera Skåne polisens användning av LUPP-metoden, används även här Kruskal-Wallis test för att analysera prognoskvalitén. Testet utförs med fem grupper representerade av varje enskilt polisområde. Syftet med denna prövning är att se om det finns några regionala skillnader med avseende på den observerade prognoskvalitén. Är något polisområde mer lyckosamt i sitt användande av LUPP än de övriga? Följande hypotes formuleras:

$H_0$ : Prognoskvalitén är lika för samtliga polisområden

$H_1$ : Prognoskvalitén skiljer sig åt för något av polisområdena

### 5. Redovisning av resultat

#### 5.1 Undersökning av prognoskvalitén - samtliga polisområden

$H_0$ : Prognoskvalitén är lika för samtliga prognosmetoder

$H_1$ : Prognoskvalitén skiljer sig åt för någon av metoderna

Resultatet för testet (tabell 5.1) visar att  $H_0$  kan accepteras ty  $P= 0,407$ . Man kan därmed konstatera att de skillnader som finns med avseende på prognoskvalitén, följaktligen inte kan förklaras av valet av prognosmetod. Dock uppvisar LUPP-metoden klart högst genomsnittligt prognosfel (84,83) och är med den utgångspunkten sämsta metod. Metoderna Tidsserieuppdelning och ARIMA får likvärdigt resultat, där dock metoden för Tidsserieuppdelning uppvisar något lägre medelvärde. Ur ett kvalitetsperspektiv kan dessa två metoder rimligtvis vara att betrakta som likvärdiga.

Tabell 5.1

Source	DF	SS	MS	F	P
Factor	2	8321	4161	0,90	0,407
Error	177	815139	4605		
Total	179	823460			

Metod	N	Mean	StDev
TIDS.UPPD.	60	68,97	62,87
LUPP	60	84,83	79,00
ARIMA	60	72,52	60,18

## 5.2 Undersökning av prognoskvalitén - enskilt polisområde

$H_0$ : Prognoskvalitén är lika för samtliga prognosmetoder

$H_1$ : Prognoskvalitén skiljer sig åt för någon av metoderna

Resultat för Kruskal-Wallis test visar att  $H_0$  accepteras för samtliga polisområden (tabell 5.2). Testet ger följaktligen inget stöd för ett antagande om en eventuell bästa prognosmetod för respektive polisområde.

Tabell 5.2

Polisområde	Testvärde	Kritiskt värde	Slutsats
Södra Skåne (POSS)	-0,152	5,991	$H_0$ accepteras
Nordvästra Skåne (PONV)	2,896	5,991	$H_0$ accepteras
Mellersta Skåne (POMS)	1,342	5,991	$H_0$ accepteras
Malmö (POMA)	0,137	5,991	$H_0$ accepteras
Norra Skåne (PONS)	0,172	5,991	$H_0$ accepteras

Testvärden för POSS, POMS och PONS är justerade för ”ties”.

## 5.3 Undersökning av prognoskvalitén med avseende på LUPP

$H_0$ : Prognoskvalitén är lika för samtliga polisområden

$H_1$ : Prognoskvalitén skiljer sig åt för något av polisområdena

Resultatet ger ett starkt stöd för mothypotesen ty testvärdet  $T: 17,20 > 9,488 \chi^2_{4,0.05}$   
(testvärdet justerat för ”ties”)

Man kan alltså konstatera en signifikant skillnad mellan de olika polisområdena med avseende på prognoskvalitén. I bilagan på sidan 44 ser man att polisområde PONS och polisområde PONV, har lägst respektive högst summerat rangtal. Detta indikerar då på att dessa polisområden är de som har den bästa, respektive den sämsta kvalitén på sina LUPP-prognoser. För att se hur relationen ser ut mellan enskilda polisområden, utfördes också parvisa tester. Olikheterna prövades enligt formel vid punkt 4.12.2 och gav följande resultat:

$$\begin{aligned}
 PONS_{\bar{R}_i} - POMA_{\bar{R}_j} &= -10,00 < 12,47 \\
 PONS_{\bar{R}_i} - POMS_{\bar{R}_j} &= -3,960 < 12,47 \\
 PONS_{\bar{R}_i} - POSS_{\bar{R}_j} &= -7,040 < 12,47 \\
 PONS_{\bar{R}_i} - PONV_{\bar{R}_j} &= -27,96 < 12,47
 \end{aligned}$$

---


$$\begin{aligned}
 PONV_{\bar{R}_i} - POMA_{\bar{R}_j} &= 17,95 > 12,47 \\
 PONV_{\bar{R}_i} - POMS_{\bar{R}_j} &= 24,00 > 12,47 \\
 PONV_{\bar{R}_i} - POSS_{\bar{R}_j} &= 20,86 > 12,47
 \end{aligned}$$

Resultatet visar på att det finns statistiskt säkerställda olikheter mellan samtliga kombinationer baserade på PONS respektive PONV. Vi kan därmed definitivt slå fast att det finns statistiskt säkerställda skillnader med avseende på LUPP-prognosernas kvalitet. Bäst fungerar LUPP således för polisen i norra Skåne (PONS) och sämst för polisområde nordvästra Skåne (PONV).

För att även pröva hypotesen utifrån genomsnittligt prognosfel, utfördes en ensidig variansanalys. Resultatet för denna prövning (tabell 5.3) är analog med Kruskal-Wallis test, och visar på signifikanta skillnader mellan de olika polisområdenas prognoskvalité. Vi kan således även här förkasta nollhypotesen och därmed konstatera att den undersökta faktorn (polisområde), har signifikant betydelse för prognosernas kvalitet. Resultatet skall dock tolkas försiktigt då antalet replikat för varje nivå är litet. Testet visar dock att polisområde PONS och PONV, har lägst respektive högst genomsnittligt prognosfel. Detta stödjer slutsatsen om att dessa områden har bäst, respektive sämst kvalitet på sina LUPP-prognoser.

Tabell 5.3

Source	DF	SS	MS	F	P
Polisomr	4	141846	35462	8,61	0,000
Error	55	226408	4117		
Total	59	368254			

Polisområde	N	Mean	StDev
POMA	12	96,83	102,34
POMS	12	50,08	26,93
PONS	12	47,50	41,38
PONV	12	175,17	84,03
POSS	12	54,58	24,69

## 6. Slutdiskussion

Sammanfattningsvis kan man alltså konstatera att det statistiskt inte går att säkerställa några skillnader mellan de olika prognosmetoderna Box-Jenkins ARIMA, Tidsserieuppdelning och polisens egen modell LUPP. Av resultatet kan man dock dra den slutsatsen att polisens modell LUPP, är att betrakta som sämre än de övriga. Detta antagande grundas på att denna metod uppvisade det klart högsta genomsnittliga prognosfelet vid analys av samtliga polisområden ackumulerat (punkt 5.1). Vid de icke-parametriska testerna för enskilda polisområden (punkt 5.2), kan man även här finna belägg för antagandet om LUPP som sämsta metod. Vid de fem tester som utfördes för respektive polisområde, uppvisade LUPP-modellen högst rangtal i fyra av dessa. De övriga metoderna i undersökningen visade sig vara likvärdiga överlag.

För att trots allt fastställa även en bästa modell, så kan metoden för Tidsserieuppdelning vara ett bra val. Denna metod är betydligt mer användarvänlig i jämförelse med Box-Jenkins ARIMA, och är av den anledningen intuitivt att föredra. Metoden har dock den svagheten att den anses vara bäst för stabila tidsserier. Detta faktum har på förhand upplevts som en nackdel, då tidsserierna för bilinbrott snarare visade på hög fluktuation än på ett stabilt mönster. Metodens karakteristik har emellertid inte varit någon nackdel i denna studie då resultatet visade sig vara analogt med ARIMA-modellerna, vilka anses vara lämpliga för just rörliga tidsserier. Att metoden för Tidsserieuppdelning visade goda resultat vid denna undersökning, kan troligen förklaras med att den tidsserie som undersökts, visade sig vara relativt jämn. Ett faktum som styrker detta antagandet är bland annat att prognoskvalitén var högre då trendkomponenten utlämnades. Av det följer också en intressant parentes rörande den allmänt rådande uppfattningen om en ökande vardagsbrottslighet. Eftersom ingen trend kunde konstateras, kan man följaktligen dra slutsatsen att denna uppfattning inte verkar stämma för brottstypen bilinbrott. Sammanfattningsvis kan således metoden för Tidsserieuppdelning vara den metod som bäst lämpar sig att prognostisera antalet framtida bilinbrott för Skåne län.

Med avseende på LUPP-prognosernas tillkortakommande, kan detta ha sin grund att de endast baseras på median värdet av observationer tre år tillbaka i tiden. Detta innebär att jämna tidsserier är en förutsättning för bra prognoser. Skulle det exempelvis inträffa en period av några års konstant brottsökning, tar det långt tid innan LUPP-prognoserna speglar utfallet. Denna eftersläpning har sin största orsak just i användande av medianvärdet. Om exempelvis antalet brott ökar tre år i rad för att sedan stabilisera sig kring en hög nivå, tar det två

årscykler innan den högre prognosen slår igenom. För att på ett framgångsrikt sätt prognostisera den framtida brottsligheten, behöver instrumenten vara betydligt mer finkänsliga. Denna studie har visat att bilinbrotten kraftigt fluktuerar över tiden, ett faktum som rimmar illa med de förutsättningar som krävs för goda LUPP-prognoser.

Vid undersökningen huruvida LUPP-modellen fungerar bättre för något enskilt polisområde, kunde emellertid statistiskt säkerställda skillnader konstateras. Detta test visar att polisområde norra Skåne (PONS) har den bästa prognoskvalitén, medan polisen nordvästra Skåne (PONV) uppvisar den sämsta kvalitén på sina LUPP-prognoser. Anledningen till detta torde helt enkelt vara att brottsnivån för område PONS förmodligen är mer stabil. LUPP-metoden är inte lämpad för tidsserier som fluktuerar, och är av den orsaken även att betrakta som ett otidsenligt verktyg.

## 7. Summary in English

The purpose of this report is to compare the forecasting model used by the Swedish police to more scientific methods used by statisticians. Is the LUPP-method considered to be competitive in terms of producing accurate forecasts? The forecasting model LUPP, is used by the Swedish police on a national basis. This method is rather uncomplicated and allows the police to establish forecasts without any prior statistical knowledge. The Lupp-modell uses a three year period where the median will be used as the forecast. For example, if a forecast is prepared for January 2006, the January actuals for 2003, 2004 and 2005 would be used. The median value will then be the forecast for 2006. This approach will take in to account any possible seasonal deviations, but the model is also sensitive for any rapid changes in the number of crimes. The model is therefore best suited for crime patterns that are considered to be stable.

The other two methods in this study are Box-Jenkins ARIMA procedure and the Decomposition method. The Arima model is convenient to use as a forecast model when the time series are rapidly changing. The Decomposition method is defined as the opposite, thus best suited for stable patterns. This investigation will be done for the police districts both together as well as separate. Apart from trying to define which forecasting model that is most convenient, there will also be a test regarding which of the five different police districts have experienced the highest forecasting quality in terms of using the LUPP-method.



The test showed that in terms of identifying the best forecasting model, this could not be statistically determined. Although the test showed that there were some differences in the quality of the forecast, this still was not strong enough to reject our hypothesis of similar forecasting quality. This result was similar when testing the districts together, as well as separate. Therefore, we have to conclude that the forecasting quality is the same for all three methods. The test figures clearly pointed out though, that the LUPP-method showed the weakest forecasts in comparison with the other two models - which were rather identical in regards to the forecast quality.

Nevertheless, to conclude the best model, the choice has to be in favour of the Decompositon method. This method is well suited for a this particular crime pattern together with the fact that this model is very simple to use in comparison with the Arima method. Although it must be stated that this conclusion could not be statistically determined.

When investigating which of the five police districts are most succesfull in their use of the LUPP-procedure, the test result allowed a rejection of the hypothesis assuming similar forecast quality. This test clearly pointed out that the police districts PONS and PONV, experience the best and the worst forecast quality.

## Bilaga

### Valda modeller Tidsserieuppdelning

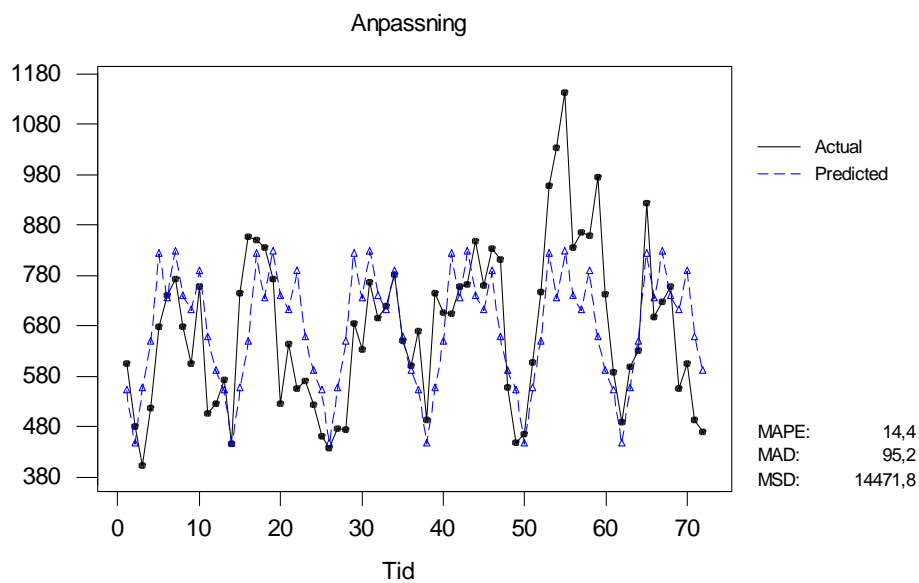
För att skapa prognosmodeller baserade på tidsserieuppdelning, har programmet Minitab använts. Prognoser för respektive polisområde har sedan skapats utifrån multiplikativa modeller utan trendkomponent. Motivet till att använda multiplikativa modeller grundar sig i karaktären för respektive tidsserie, där samtliga tidsserier uppvisar en varierande grad av icke konstant säsongsvariation.

Valet att utelämna trendkomponenten baseras på det faktum att modeller med trend, överlag gav sämre prognoskvalité. Ett förbiseende av trendkomponenten stöds också av det faktum att prognoserna endast blev något bättre när multiplikativ modell användes. Vid en tydlig trend i materialet, hade skillnaden mellan en multiplikativ och en additiv modell varit större. Trendkomponenten är dock med i modellerna men i form av ett konstant värde. Detta värde är således lika för alla  $Y_t$ . För att minimera bilagans omfattning redovisas enbart data för de 12 första observationerna. I och med detta redovisas också samtliga säsongindex ("kolumn: seasonal").

## Polisområde nordvästra Skåne (PONV)

Row	Bilinbrott PONV 98-2003	Trend	Seasonal	Detrend	Deseason	Model
1	605	672,111	0,81994	0,90015	737,862	551,088
2	479	672,111	0,66156	0,71268	724,049	444,640
3	402	672,111	0,82649	0,59812	486,396	555,491
4	517	672,111	0,96368	0,76922	536,486	647,699
5	679	672,111	1,22348	1,01025	554,976	822,312
6	740	672,111	1,09269	1,10101	677,226	734,411
7	773	672,111	1,23170	1,15011	627,587	827,841
8	677	672,111	1,09801	1,00727	616,573	737,982
9	604	672,111	1,05815	0,89866	570,805	711,197
10	757	672,111	1,17142	1,12630	646,225	787,323
11	506	672,111	0,97590	0,75285	518,494	655,916
12	524	672,111	0,87699	0,77963	597,500	589,433

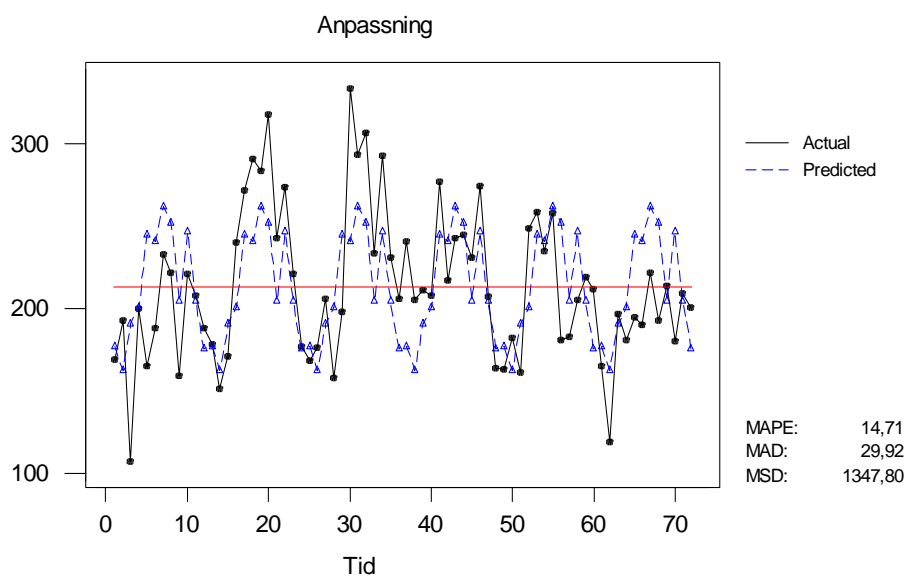
Prognoser för  $Y_t$  skapas genom att multiplicera trenden med säsongsindex. En prognos för januari 2004 blir således produkten av  $672,111 * 0,81994$ .



## Polisområde södra Skåne (POSS)

Row	Bilinbrott 98-2003	Trend	Seasonal	Detrend	Deseason	Model
1	169	213,542	0,82921	0,79141	203,808	177,072
2	193	213,542	0,75992	0,90380	253,973	162,275
3	107	213,542	0,89269	0,50107	119,863	190,626
4	200	213,542	0,93853	0,93659	213,100	200,415
5	165	213,542	1,14823	0,77268	143,699	245,195
6	188	213,542	1,12848	0,88039	166,596	240,977
7	233	213,542	1,22657	1,09112	189,960	261,925
8	222	213,542	1,18117	1,03961	187,949	252,229
9	159	213,542	0,95896	0,74459	165,805	204,777
10	221	213,542	1,15483	1,03493	191,370	246,605
11	208	213,542	0,95786	0,97405	217,150	204,544
12	188	213,542	0,82355	0,88039	228,281	175,862

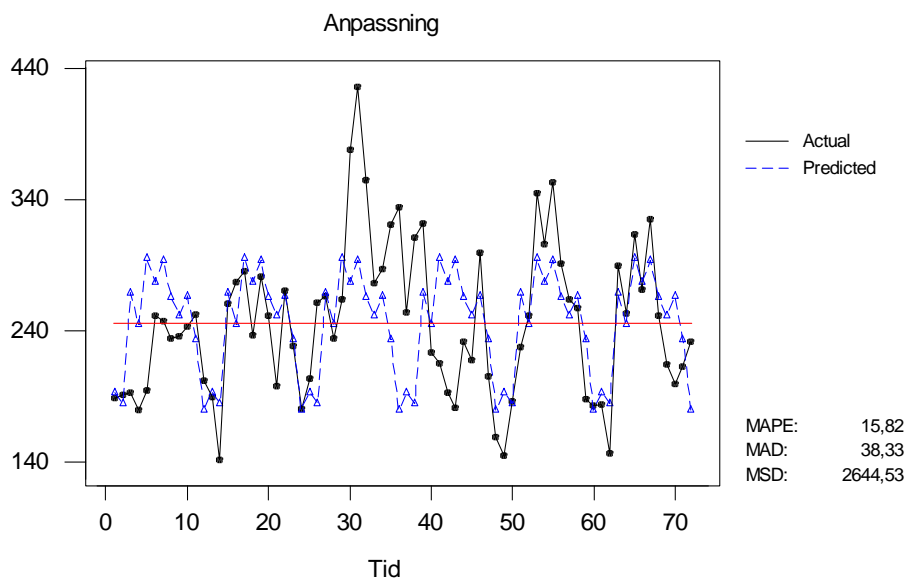
Prognoser skapas genom:  $213,542 * \text{Säsongindex för given säsong.}$



## Polisområde mellersta Skåne (POMS)

Row	Bilinbrott 98-2003	Trend	Seasonal	Detrend	Deseason	Model
1	188	245,764	0,78318	0,76496	240,046	192,478
2	191	245,764	0,75005	0,77717	254,650	184,335
3	192	245,764	1,09288	0,78124	175,682	268,591
4	179	245,764	0,99332	0,72834	180,204	244,122
5	194	245,764	1,19966	0,78938	161,712	294,834
6	251	245,764	1,12698	1,02131	222,719	276,972
7	247	245,764	1,19380	1,00503	206,902	293,393
8	234	245,764	1,07841	0,95213	216,986	265,035
9	235	245,764	1,02284	0,95620	229,753	251,376
10	243	245,764	1,08137	0,98875	224,714	265,763
11	252	245,764	0,94917	1,02537	265,494	233,272

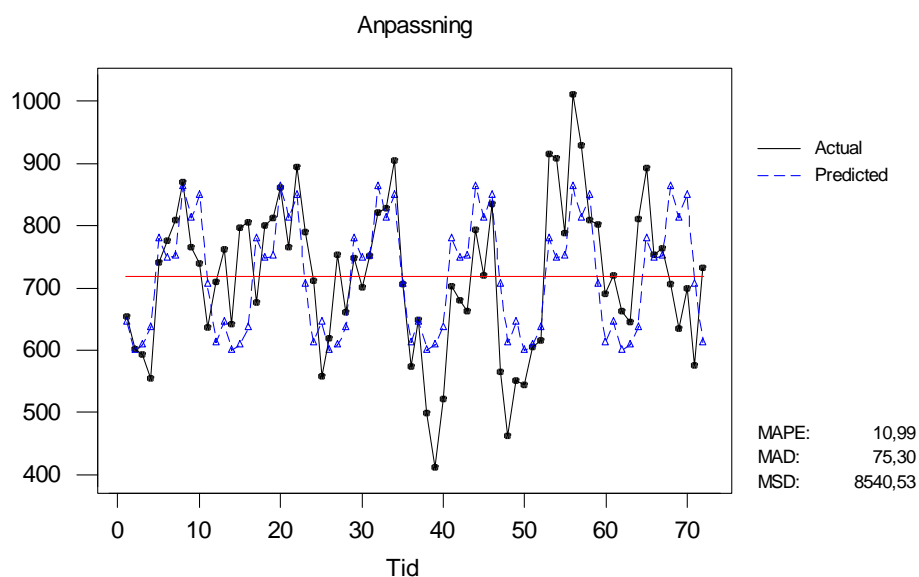
Prognoser skapas genom:  $245,764 * \text{Säsongindex för given säsong}$ .



## Polisområde Malmö (POMA)

Row	Bilinbrott POMA 98-2003	TREN2	SEAS2	Detrend	DESE2	Model
1	654	717,653	0,89881	0,91130	727,630	645,032
2	602	717,653	0,83500	0,83885	720,961	599,238
3	592	717,653	0,84654	0,82491	699,315	607,524
4	554	717,653	0,88547	0,77196	625,655	635,462
5	741	717,653	1,08614	1,03253	682,230	779,474
6	776	717,653	1,04319	1,08130	743,872	748,648
7	809	717,653	1,04650	1,12729	773,056	751,021
8	870	717,653	1,20382	1,21229	722,699	863,925
9	765	717,653	1,13298	1,06598	675,208	813,089
10	739	717,653	1,18442	1,02975	623,933	850,003
11	637	717,653	0,98384	0,88762	647,460	706,059
12	709	717,653	0,85328	0,98794	830,911	612,359

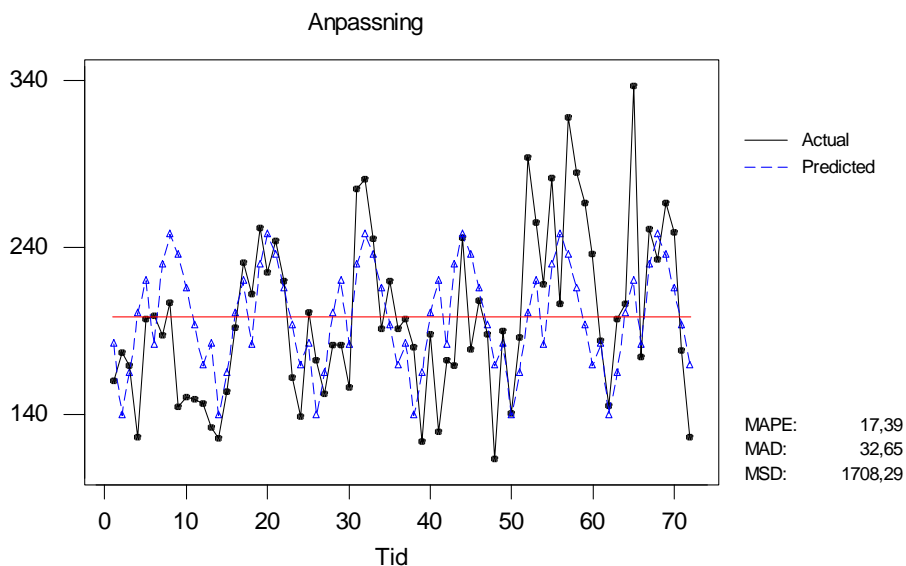
Prognoser skapas genom:  $717,653 * \text{Säsongindex för given säsong.}$



## Polisområde norra Skåne (PONS)

Row	Bilinbrott 98-2003	Trend	Seasonal	Detrend	Deseason	Model
1	160	198,042	0,91931	0,80791	174,043	182,063
2	177	198,042	0,70279	0,89375	251,855	139,181
3	169	198,042	0,82926	0,85336	203,796	164,228
4	126	198,042	1,01027	0,63623	124,719	200,075
5	197	198,042	1,11019	0,99474	177,448	219,863
6	199	198,042	0,91505	1,00484	217,474	181,219
7	187	198,042	1,15960	0,94425	161,263	229,649
8	207	198,042	1,25004	1,04523	165,594	247,561
9	144	198,042	1,18884	0,72712	121,127	235,440
10	150	198,042	1,08536	0,75742	138,204	214,946
11	149	198,042	0,97575	0,75237	152,702	193,240
12	146	198,042	0,85354	0,73722	171,052	169,037

Prognoser skapas genom:  $198,042 * \text{Säsongindex för given säsong.}$



## Valda ARIMA-modeller

### Polisområde nordvästra Skåne (PONV)

Modell:

Final Estimates of Parameters

Type	Coef	SE Coef	T	P
MA 1	-0,5607	0,1277	-4,39	0,000
MA 2	-0,4625	0,1335	-3,46	0,001
MA 3	-0,4015	0,1277	-3,14	0,003
SMA 12	0,7980	0,1447	5,52	0,000

Differencing: 0 regular, 1 seasonal of order 12

Number of observations: Original series 72, after differencing 60

Residuals: SS = 195,921 (backforecasts excluded)  
MS = 3,499 DF = 56

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

Lag	12	24	36	48
Chi-Square	11,3	23,7	41,3	59,6
DF	8	20	32	44
P-Value	0,187	0,255	0,126	0,058

Slutgiltig modell PONV:

$Z_t = y_t - y_{t-L} + y_{t-L-1} + y_{t-L-2} + y_{t-L-3}$  där prognoser för  $y_t$  skapas genom:

$$y_t = y_{t-12} - y_{t-13} - y_{t-14} - y_{t-15} - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \theta_3 a_{t-3} - \theta_{1,12} a_{t-12} + \theta_1 \theta_{1,12} a_{t-13} + \theta_2 \theta_{1,12} a_{t-14} + \theta_3 \theta_{1,12} a_{t-15} + a_t$$

### Polisområde södra Skåne (POSS)

Modell:

Final Estimates of Parameters

Type	Coef	SE Coef	T	P
AR 1	0,4670	0,1100	4,24	0,000
MA 1	0,9843	0,0157	62,77	0,000
SMA 12	-0,3486	0,1221	-2,85	0,006

Differencing: 1 regular difference

Number of observations: Original series 72, after differencing 71

Residuals: SS = 95435,6 (backforecasts excluded)  
MS = 1403,5 DF = 68

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

Lag	12	24	36	48
Chi-Square	5,2	16,4	35,1	44,2
DF	9	21	33	45
P-Value	0,817	0,745	0,369	0,507

Slutgiltig modell POSS:

$Z_t = y_t - y_{t-L}$  där prognoser för  $y_t$  skapas genom:

$$y_t = y_{t-12} + \Phi_1 Z_{t-1} - \theta_1 a_{t-1} - \theta_{1,12} a_{t-12} + a_t$$



## Polisområde mellersta Skåne (POMS)

### Modell

#### Final Estimates of Parameters

Type	Coef	SE Coef	T	P
AR 1	0,5588	0,1130	4,94	0,000
SAR 12	-1,0317	0,1101	-9,37	0,000
SAR 24	-0,8708	0,1111	-7,84	0,000

Differencing: 0 regular, 1 seasonal of order 12

Number of observations: Original series 72, after differencing 60

Residuals: SS = 1,74207 (backforecasts excluded)  
MS = 0,03056 DF = 57

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

Lag	12	24	36	48
Chi-Square	7,9	19,9	28,2	30,9
DF	9	21	33	45
P-Value	0,546	0,526	0,707	0,946

### Slutgiltig modell POMS:

$Z_t = y_t - y_{t-L} + y_{t-L-1} + y_{t-2L-1}$  där prognoser för  $y_t$  skapas genom:

$$y_t = y_{t-12} - y_{t-13} - y_{t-25} + \Phi_1 Z_{t-1} + \Phi_{1,12} Z_{t-12} + \Phi_{2,12} Z_{t-24} + \Phi_1 \Phi_{1,12} Z_{t-13} + \Phi_1 \Phi_{2,12} Z_{t-25} + a_t$$

## Polisområde Malmö (POMA)

### Modell:

#### Final Estimates of Parameters

Type	Coef	SE Coef	T	P
AR 1	0,5449	0,1109	4,91	0,000
SMA 12	0,7971	0,1434	5,56	0,000

Differencing: 0 regular, 1 seasonal of order 12

Number of observations: Original series 72, after differencing 60

Residuals: SS = 176,115 (backforecasts excluded)  
MS = 3,036 DF = 58

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

Lag	12	24	36	48
Chi-Square	13,8	33,3	46,7	69,0
DF	10	22	34	46
P-Value	0,184	0,057	0,072	0,016

### Slutgiltig modell POMA:

$Z_t = y_t - y_{t-L}$  där prognoser för  $y_t$  skapas genom:

$$y_t = y_{t-12} + \Phi_1 Z_{t-1} - \theta_{1,12} a_{t-12} + a_t$$

## Polisområde norra Skåne (PONS)

Modell:

### Final Estimates of Parameters

Type	Coef	SE Coef	T	P
MA 1	-0,2311	0,1197	-1,93	0,059
MA 2	-0,4341	0,1206	-3,60	0,001
SMA 12	0,7441	0,1374	5,41	0,000

Differencing: 0 regular, 1 seasonal of order 12

Number of observations: Original series 72, after differencing 60

Residuals: SS = 98740,8 (backforecasts excluded)  
MS = 1732,3 DF = 57

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

Lag	12	24	36	48
Chi-Square	10,1	16,4	24,0	28,9
DF	9	21	33	45
P-Value	0,346	0,745	0,873	0,970

Slutgiltig modell PONS:

$Z_t = y_t - y_{t-L} + y_{t-L-1} + y_{t-L-2}$                       där prognoser för  $y_t$  skapas genom:

$y_t = y_{t-12} - y_{t-13} - y_{t-14} - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \theta_{1,12} a_{t-12} + \theta_1 \theta_{1,12} a_{t-13} + \theta_2 \theta_{1,12} a_{t-14} + a_t$

Antal bilinbrott 1998-2004

## Polisområde nordvästra Skåne (PONV)

	Jan	Feb	Mar	Apr	Maj	Jun	Jul	Aug	Sep	Okt	Nov	Dec
1998	605	479	402	517	679	740	773	677	604	757	506	524
1999	573	444	744	858	851	835	772	525	643	554	571	523
2000	460	436	475	473	684	632	766	695	720	782	649	600
2001	670	492	746	707	704	757	763	849	760	834	811	558
2002	448	465	607	747	958	1 034	1 144	835	866	859	975	743
2003	587	489	598	630	925	697	728	758	554	604	493	469
2004	406	425	503	511	667	721	657	579	596	601	500	365

## Polisområde södra Skåne (POSS)

	Jan	Feb	Mar	Apr	Maj	Jun	Jul	Aug	Sep	Okt	Nov	Dec
1998	169	193	107	200	165	188	233	222	159	221	208	188
1999	178	151	171	240	272	291	284	318	243	274	221	177
2000	168	176	206	158	198	334	294	307	234	293	231	206
2001	241	205	211	208	277	217	243	245	231	275	207	164
2002	163	182	161	249	259	235	258	181	183	205	219	212
2003	165	119	197	181	195	190	222	193	214	180	209	201
2004	251	248	163	181	205	275	288	305	274	247	238	243

## Polisområde mellersta Skåne (POMS)

	Jan	Feb	Mar	Apr	Maj	Jun	Jul	Aug	Sep	Okt	Nov	Dec
1998	188	191	192	179	194	251	247	234	235	243	252	201
1999	189	141	260	277	285	236	281	251	197	270	228	180
2000	203	261	266	234	264	378	426	355	276	287	321	334
2001	254	311	322	223	215	192	181	231	217	299	205	158
2002	144	186	227	251	345	306	353	291	264	257	187	182
2003	183	146	289	253	313	271	325	251	214	199	212	231
2004	236	225	247	229	263	253	214	227	261	345	255	242

## Polisområde Malmö (POMA)

	Jan	Feb	Mar	Apr	Maj	Jun	Jul	Aug	Sep	Okt	Nov	Dec
1998	654	602	592	554	741	776	809	870	765	739	637	709
1999	763	642	797	806	676	800	813	862	766	896	790	712
2000	557	619	754	660	749	701	752	822	829	906	707	573
2001	649	498	411	520	703	680	663	794	720	835	564	461
2002	550	544	605	615	917	910	788	1 012	930	810	803	690
2003	720	662	645	811	894	754	764	707	635	700	575	732
2004	665	715	663	580	538	757	711	618	717	702	715	733

## Polisområde norra Skåne (PONS)

	Jan	Feb	Mar	Apr	Maj	Jun	Jul	Aug	Sep	Okt	Nov	Dec
1998	160	177	169	126	197	199	187	207	144	150	149	146
1999	132	125	153	192	231	212	252	225	244	220	162	138
2000	201	172	152	181	181	156	275	281	245	191	220	191
2001	197	180	123	188	129	172	169	246	179	208	188	113
2002	190	140	186	294	255	218	282	206	318	285	267	236
2003	184	145	197	206	337	174	251	233	267	249	178	126
2004	144	117	206	175	181	339	259	206	203	207	211	168

Kruskal-Wallis test (punkt 5.2)

## Polisområde nordvästra Skåne (PONV)

LUPP		ARIMA		TIDS.UPPD.			
Prognosfel	Rang	Prognosfel	Rang	Prognosfel	Rang		
181		25		136	13,5	145	16
64		6		43	4	19	2
104		9		93	8	52	5
196		30		149	17	136	13,5
258		35		157	20	155	18,5
36		3		81	7	13	1
106		10		195	29	170	24
256		34		162	22	158	21
164		23		107	11	115	12
233		33		139	15	186	26
311		36		187	27	155	18,5
193		28		206	31	224	32
<hr/>							
R <sub>ij</sub>		272			204,5		189,5
n <sub>ij</sub>		12			12		12
N = 36							

## Polisområde södra Skåne (POSS)

LUPP		ARIMA		TIDS.UPPD.			
Prognosfel	Rang	Prognosfel	Rang	Prognosfel	Rang		
86		33,5		45	17,5	74	30
66		27		65	26	86	33,5
34		10,5		51	20	27	5,5
27		5,5		16	3	19	4
54		22		1	1,5	40	14
58		24		75	31	35	12
45		17,5		78	32	27	5,5
112		36		96	35	53	21

	60	25	57	23	70	29
	42	15,5	45	17,5	1	1,5
	29	8,5	29	8,5	34	10,5
	42	15,5	38	13	68	28
$R_{ij}$		240,5		228		194,5
$n_{ij}$		12		12		12
$N = 36$						

## Polisområde mellersta Skåne (POMS)

LUPP		ARIMA		TIDS.UPPD.		
Prognosfel	Rang	Prognosfel	Rang	Prognosfel	Rang	
53	26	2		1	44	20,5
39	17	68		31	41	18
42	19	58		28	21	8
22	9,5	3		2	15	5
50	23,5	34		14	31	13
18	7	49		22	23	11
111	36	17		6	79	32
24	12	12		4	38	16
44	20,5	37		15	10	3
88	35	50		23,5	80	33
50	23,5	54		27	22	9,5
60	29	83		35	64	30
$R_{ij}$		258		208,5		199
$n_{ij}$		12		12		12
$N = 36$						

## Polisområde Malmö (POMA)

LUPP		ARIMA		TIDS.UPPD.		
Prognosfel	Rang	Prognosfel	Rang	Prognosfel	Rang	
16	5	29		8	20	6
171	30	127		27	116	25
58	18	44		13	56	17
35	9	78		20	55	16
356	36	264		35	241	33
3	1,5	25		7	9	3,5
53	15	52		14	40	11
176	31	233		32	245	34
3	1,5	62		19	96	22
108	24	100		23	148	29
140	28	38		10	9	3,5
43	12	92		21	121	26
$R_{ij}$		211		229		226
$n_{ij}$		12		12		12
$N = 36$						



Polisområden inom Skåne län

Nordvästra skåne (PONV)

Innefattar följande kommuner:

Helsingborg  
 Bjuv  
 Höganäs  
 Landskrona  
 Svalöv  
 Ängelholm  
 Båstad  
 Klippan  
 Åstorp  
 Örkeljunga

Malmö (POMA)

Innefattar följande kommuner:

Malmö  
 Burlöv

Nordöstra skåne (PONS)

Innefattar följande kommuner:

Kristianstad  
 Ö Göinge  
 Bromölla  
 Hässleholm  
 Osby  
 Perstorp

Mellersta skåne (POMS)

Innefattar följande kommuner:

Lund  
 Staffanstorps  
 Kävlinge  
 Eslöv  
 Höör  
 Hörby  
 Lomma

Södra skåne (POSS)

Innefattar följande kommuner:

Ystad  
 Skurup  
 Sjöbo  
 Svedala  
 Vellinge  
 Trelleborg  
 Simrishamn  
 Tomelilla

## Källförtäckning

### Publicerade källor:

Ball, Cullen, Lilly. *Criminological Theory Context and Consequenses* (2001). Sage Publications

Bergström, Reinhold. *Prognoser i praktiken* (1982). LiberFörlag

Boweman, o'Connell. *Forecasting and Time Series* (1993). Duxbury Classic series

Conover, W.J. *Practical Nonparametric Statistics* (1999). John Wiley & Sons, Inc.

Flemming. C. Michael. Nellis, Joseph, G. *Applied Statistics* (2000). Thomson Learning

Montgomery, D.M. *Design and Analysis of Experiments* (2001). John Wiley & Sons, Inc.

Brottsförebyggande rådet. *Kodning av brott, BRÅ rapport, version 2, REV:2005-10-03*

LUPP, *En lokal uppföljnings- & prognosprocedur i problemorienterat arbete mot brott.*  
BRÅ dokument: 91-38-31887-3:2001

### Elektroniska källor:

www.[minitab.com](http://www.minitab.com)