



LUNDS UNIVERSITET
Ekonomihögskolan

ARIMA-modellering av anlagda bilbränder i Sverige

Amelie Persson

Kandidatuppsats i statistik, 15hp
Statistiska institutionen
Lunds universitet
Handledare: Jonas Wallin
HT 2018

Abstract

An ongoing trend the last couple of years has been the intense reporting of an increased amount of deliberate vehicle fires in Sweden. As the overall assumption in society is that the number of deliberate vehicle fires is steadily increasing, a statistic model trying to investigate this was made. In 1998, the share of deliberate car fires were estimated to 12%, the same number in 2015 was 38%, indicating a growing problem. In addition, these types of criminal behaviour is characterized by a high number of criminals not getting convicted since there seldom is neither a lot of witnesses nor technical evidence.

By dividing data collected from the Swedish Civil Contingencies Agency into three subcategories, namely large cities, medium-sized towns and smaller towns, a time series analysis was conducted. To further concretize the difference between these subcategories, the data was collected as a relative number, the number of deliberate car fires per 1000 residents. All available data in the database of the Swedish Civil Contingencies Agency were used, meaning the analysis contained data between the years 1998 and 2017.

The analysis was made by fitting ARIMA-models, Autoregressive Integrated Moving Average-models, to the data in order to enable the creation of forecasts as well as prediction intervals. By removing a subset of the last 10% of the data before the analysis (datapoints from 2016 and 2017), a form of reference was made to observe the quality of the predictions. The prediction intervals were found to be too narrow in the case of large cities and smaller towns when compared to the subset. A possible reason for this may be that the real observations made between 2016 and 2017 contained sporadic unforeseen events, such as rebellions. The feared positive upwarding trend was confirmed as all predictions predicted a steady increase in these types of crimes.

Innehåll

1	Inledning	1
2	Teori och bakgrund	2
2.1	Bakgrund	2
2.1.1	Stationäritet	2
2.1.2	Differentiering och säsongjustering	3
2.2	AR(p) Autoregressiv process	3
2.3	MA(q) Glidande medelvärdesprocess	3
2.4	ARIMA: Autoregressiv Integrerande Glidande Medelvärdesprocess	3
2.5	Autokorrelations- och partiella autokorrelationsfunktionen	4
2.5.1	Modelltester- och informationskriterier	4
2.6	Box och Jenkins metod	5
2.7	Prediktion	5
3	Data	6
4	Beräkningar och resultat	7
4.1	Storstäder och storstadsnära kommuner	7
4.1.1	Bakgrundsinformation	7
4.1.2	Modellskapande	8
4.1.3	Prediktion	12
4.2	Större städer och kommuner nära större stad	13
4.2.1	Bakgrundsinformation	13
4.2.2	Modellskapande	14
4.2.3	Prediktion	18
4.3	Mindre städer och landsbygd	19
4.3.1	Bakgrundsinformation	19
4.3.2	Modellskapande	20
4.3.3	Prediktion	23
5	Diskussion och slutsatser	24
	Litteraturförteckning	25
	Bilagor	26

1. Inledning

Syftet med denna uppsats är att modellera och försöka prognostisera den framtida utvecklingen av anlagda bilbränder i Sverige. Genom att dela upp samhället i tre olika delar, mindre städer, större städer och storstäder, är ambitionen att analysera den framtida utvecklingsskillnaden dem emellan.

Med hjälp av räddningstjänstens insatsstatistik har Myndigheten för samhällsskydd och beredskap (MSB) sedan 1998 registrerat 19 862 anlagda personbilsbränder i Sverige, vilka kommer ligga till grund för arbetet i denna uppsats. Generellt har bilden i massmedia av antalet anlagda bilbränder de senaste åren målats upp som problematisk och ökande, vilket bland annat resulterat i ett växande allmänintresse för att denna typen av brott ska minska och utredas i större utsträckning (Kudo, 2018). År 1998 var andelen personbilsbränder som ansågs anlagda 12%, år 2015 var samma siffra 38% (MSB, 2015). Lars Förstell, presstalesperson polisregion Syd, teoretiserar kring orsaken för detta och pratar om att bilbränder genererar fler bilbränder (Nygren, 2016). Det vill säga att de som tänder eld på bilar inspirerar fler att göra samma sak. En teori som delas av Anders Ahlström, brandchef i Uppsala (Lind, 2016). Vidare problematiseras brottstypen av att väldigt få personer lagförs, 0.6% av förövarna, eftersom brotten ofta sker sent på kvällen med få vittnen och lite teknisk bevisning (Kudo, 2018).

I 2017 års upplaga av rapporten *Försäkringsbedrägerier i Sverige 2017* la Svensk Försäkring och Larmtjänst AB, båda branschorganisationer för svenska försäkringsbolag, extra fokus på en gemensam kontrollaktivitet för just bilbränder (Svensk Försäkring, 2017). Detta innebar att försäkringsbolagen la extra resurser på utredningar vad gällde granskning om ersättning efter bilbrand. Den grundläggande orsaken till detta fokus var den stora kostnaden bilbränder har för samhället liksom en misstanke om att oärliga människor i stor utsträckning försöker tjäna pengar på att begå försäkringsbedrägeri i samband med bilbränder. Vad man fann var att av 469 försäkringsutredningar så ansågs 52% (245 st) vara försök till bedrägeri, vilka beräknades uppgå till ett värde av 12.5 miljarder kronor. Med andra ord hade 52% av anmälarna, själva, eller med hjälp av andra, anlagt brand på sin egen bil med syfte att åstadkomma ekonomisk vinning. Med tanke på att endast 7% av de totala antalet skadeanmälningar granskades kan läsaren själv fundera på hur omfattade och dyr denna typ av brottslighet potentiellt är för samhället. I de fall som polis beslutade att upprätta en polisanmälan slog rapporten fast att endast 14% till slut leder till åtal (ibid.).

Den samhällsekonomiska kostnaden i samband med bilbränder kategoriserar MSB i sin statistik som direkt, indirekt eller förebyggande (MSB, 2011). De indirekta kostnaderna är kostnader av varor och tjänster som skulle kunna ha producerats om inte människor dött eller fått minskad arbetskapacitet. Direkta kostnader är de kostnader som förbrukas i samband med olyckan, det vill säga kostnader kopplade till egendomsskador och personskador men även miljöskaador, rättsväsende, vårdinstanser, räddningstjänst och försäkringsadministration (ibid.). Förebyggande kostnader består i informationspridning, utbildning och beredskap (MSB, 2014). Som med de flesta samhällsliga företeelser uppstår problem när exakta kostnader skall utarbetas. För bilbränder generellt och anlagda bilbränder specifikt, är denna typen av beräkning svår och tidskrävande vilket har lett till att några exakta siffror inte finns. En uppskattning av kostnaden för alla bränder i Sverige gjordes 2005 till 5.7 miljarder kronor, där näringslivet fick betala 82% av notan (Nationellt centrum för lärande från olyckor, 2008). Förberedande kostnader samma år beräknades uppgå till 16.7 miljarder kronor (MSB, 2014). Med utgångspunkt i MSBs statistik ses att avsiktlig brand på personbil utgör 1.3% av det totala antalet bränder, med enkel matematik resulterar detta i samhällskostnader om cirka 730 miljoner kronor för denna kategori av händelser. Dock måste poängteras att denna typ av beräkning är extremt förenklad och ämnar enbart ge en indikation på problemets omfattning.

2. Teori och bakgrund

2.1 Bakgrund

Tidsseriedata är data som samlats in med jämna mellanrum över tid (Chan och Cryer, 2008). Om värdena kan prognostiseras, förutspås, helt och hållet med hjälp av tidigare värden så sägs tidsserien vara deterministisk (Chatfield, 2011). I motsats till deterministiska tidsserier finns stokastiska, eller slumpmässiga, vilka är vanligast förekommande i verkligheten. Stokastiska tidsserier påverkas delvis eller inte alls av tidigare värden och kallas även stokastiska processer. Målet med tidsserieanalys är att försöka modellera stokastiska processer och på så vis göra lämpliga prognostiseringar av tidsseriens uppskattade framtida värden. Den stokastiska processen, $\{Y_t\}$, består endast av de observerade värdena y_t och tid, där $t=1,2,3,\dots,T$ (ibid.). Att skapa Box och Jenkins ARIMA-modeller, eller Autoregressiva Integrerade Glidande Medelvärdes-modeller, är en väl använd metod för just modellering av stokastiska processer, varför deras tillvägagångssätt kommer användas i denna uppsats.

I uppsatsen kommer också Box-Jenkins metodologi att användas som riktlinje i modellskapandet. Det var i boken *Time Series Analysis: Forecasting and Control* från 1970 som de båda statistikerna Box och Jenkins presenterade den idag välrenommerade metoden (Box et al., 2013). Metoden går ut på att beskriva och förstå de stokastiska processer som finns i en tidsserie med hjälp av AR, MA, ARMA och ARIMA-modeller. Innan en genomgång av dessa görs är det dock viktigt att beskriva det grundläggande antagandet av stationäritet som gäller hos dessa modeller (ibid.). Statistikprogrammet *R Studio* har använts under arbetet.

2.1.1 Stationäritet

Stokastiska processer kan vara strikt eller svagt stationära (Chan och Cryer, 2008). Strikt stationäritet är svårare att uppnå och svag stationäritet är det som ofta antas föreligga i tidsserieanalys, så även i denna uppsats. I en strikt stationär tidsserie Y_t är dess egenskaper oförändrade vid en tidsförskjutning. Med andra ord antas att sannolikhetsfördelningen för $Y_{t_1}, Y_{t_2}, \dots, Y_{t_m}$ ska vara densamma som sannolikhetsfördelningen för $Y_{t_1-k}, Y_{t_2-k}, \dots, Y_{t_m-k}$, där k representerar tidsförskjutningen (ibid.).

Vid svag stationäritet gäller att väntevärdet och variansen är konstant vid alla värden för t , eftersom processen, för alla värden på Y_t , är identiskt fördelad (ibid.). Samtidigt ska autokovariansen, alltså samvariationen mellan Y_t och Y_{t+k} , enbart bero på storleken på tidsförskjutningen k (en mer ingående definition för autokovarians återfinns under den senare rubriken *Autokorrelations- och partiella autokorrelationsfunktionen*). Eftersom varians är en speciell typ av autokovarians där $k = 0$ medför kraven för svag stationäritet att

$$\begin{array}{ll} E[Y_t] = \mu & \text{konstant för alla } t \\ Cov[Y_t, Y_{t+k}] = \gamma_k & \text{beror på storleken av tidsförskjutningen, } k \\ Var[Y_t] = \gamma_0 & \text{konstant för alla } t \end{array}$$

(Chatfield, 2011). I de allra flesta situationer där tidsserier skall analyseras finns någon eller några slags trender, säsongvariationer eller cykliska variationer som gör att kravet för stationäritet inte uppfylls. Vid dessa fall är det viktigt att på olika sätt modifiera tidsserien så att processen blir stationär (ibid.). I dagsläget finns en mängd olika metoder för att testa stationäritet och vi kommer i denna uppsats att använda Dickey-Fullers test för enhetsrot. Testets nollhypotes, att en enhetsrot existerar, förkastas om datamaterialet är stationärt. Om datamaterialet i sin ursprungliga form är icke-stationärt men efter en säsongsdifferentiering blir stationärt så tolkas det som att det finns en (1) enhetsrot. Ett värde under den kritiska 5 procentgränsen, -1.96 , indikerar att materialet är stationärt (Chan och Cryer, 2008).

2.1.2 Differentiering och säsongjustering

Om en process uppvisar säsongvariationer, alltså mönster som uppträder med jämna mellanrum under ett år, såsom varje kvartal, månad eller vecka, är det nödvändigt att ta bort dem för att uppnå stationaritet (Chan och Cryer, 2008). Så kallad säsongjustering går ut på att ta bort säsongvariationer och proceduren går ut på att subtrahera säsongkomponenten från originaldata, där säsongkomponenten består av medelvärdet för den säsongvariation som önskas elimineras (ibid.).

$$\nabla_s Y_t = Y_t - Y_{t-s}$$

där ∇ står för differentiering och s står för säsongperiodiciteten, vilken för månadsdata är 12. Proceduren för månadsdata hade varit att först beräkna ett sammantaget medelvärde för alla januari, februari och så vidare, som sedan subtraheras från varje enskilt originalvärde (ibid.).

På ett liknande sätt justeras processer som uppvisar nedåt- eller uppåtgående trender genom så kallad differentiering, alltså att föregående observation subtraheras från nuvarande observation (ibid.).

$$\nabla Y_t = Y_t - Y_{t-1}$$

En trend är, till skillnad från en säsongvariation, en period av längre trend åt ett och samma håll (Chatfield, 2011). Det är viktigt att komma ihåg att en process kan uppvisa både säsongvariation och trend och alltså kan flera metoder för differentiering behöva användas (ibid.).

Cykliska variationer är mönster som sträcker sig längre än ett år, exempelvis längre konjunkturcykler, vilka är desto svårare att komma åt med hjälp av differentiering och säsongjustering (ibid.). Resttermer, eller feltermer, är de som kvarstår när cykliska variationer, säsongvariationer och trender tagits bort (ibid.).

2.2 AR(p) Autoregressiv process

För en autoregressiv process beror variabeln linjärt på sina egna föregående värden och en stokastisk term, med andra ord är processen en typ av stokastisk differensekvation och beskrivs nedan (Chan och Cryer, 2008)

$$\text{AR}(p) : Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + e_t$$

p symboliserar antalet parametrar processen beror på

Där den stokastiska termen, e_t , är en helt slumpmässig process med väntevärde 0 och konstant varians (Chatfield, 2011).

2.3 MA(q) Glidande medelvärdesprocess

En tidsserie följer en glidande medelvärdesprocess av ordning q om den är uppbyggd som en viktad linjär summa av sina egna q viktade föregående slumpmässiga värden (ibid.).

$$\text{MA}(q) : Y_t = e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \dots - \theta_q e_{t-q}$$

q symboliserar antalet parametrar processen beror på

Variabeln beror linjärt på nuvarande och tidigare värden av en stokastisk, icke-förutsägbar term e_t (Shumway och Stoffer, 2011). Den stokastiska termen e_t är även den en helt slumpmässig process med väntevärde 0 och konstant varians (Chatfield, 2011).

2.4 ARIMA: Autoregressiv Integrerande Glidande Medelvärdesprocess

Genom att kombinera autoregressiva processer och glidande medelvärdesprocesser så får man ARMA(p,q) modeller (ibid.).

$$\text{ARMA}(p,q) : Y_t = \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} + e_t - \theta_1 e_{t-1} - \dots - \theta_q e_{t-q}$$

ARIMA(p,d,q) är en typ av ARMA modeller som behandlar tidsserier som blivit differentierade ett visst antal gånger, där I står för Integrated och d är antalet gånger tidsserien blivit differentierad för att uppnå stationaritet.

$$\text{ARIMA}(p,1,q) : Y_t - Y_{t-1} = \phi_1(Y_{t-1} - Y_{t-2}) + \phi_2(Y_{t-2} - Y_{t-3}) \dots + \phi_p(Y_{t-p} - Y_{t-p-1}) + e_t - \theta_1 e_{t-1} - \dots - \theta_q e_{t-q}$$

1 innebär att tidsserien differentierats en gång

(Chan and Cryer, 2008). Vilket också kan skrivas som

$$Y_t = (1 + \phi_1)Y_{t-1} + (\phi_2 - \phi_1)Y_{t-2} + \dots + (\phi_p - \phi_{p-1})Y_{t-p} + \phi_p Y_{t-p-1} + e_t - \theta_1 e_{t-1} - \dots - \theta_q e_{t-q}$$

En SARIMA är, i likhet med ARIMA, en form av ARMA-modeller vilken speciellt tar säsongsdifferentierade tidsserier i beaktande, där S står för säsongsvariation. I modellen, som skrivs SARIMA(p,d,q)(P,D,Q)_s, inkluderas även P, Q och D som är de säsongsvarierade versionerna av den autoregressiva ordningen (P), glidande medelvärdesordningen (Q) och differentieringen (D). Indexet s , i slutet av modellen, symboliserar längden för en säsong, exempelvis 12 för månadsindelade data med säsongsvariationer varje år (Chatfield, 2011).

2.5 Autokorrelations- och partiella autokorrelationsfunktionen

Att studera autokorrelationsfunktioner (ACF) och partiella autokorrelationsfunktioner (PACF) är metoder för att avgöra vilken ARIMA-modell som bäst passar givna data (Bisgaard och Murat, 2011). För att förstå hur en ACF fungerar är det nödvändigt att förstå hur autokovarians och autokorrelation mellan två variabler i samma tidsserie ser ut. Autokovarians skrivs

$$\text{Cov}(Y_{t+k}, Y_t) = E[(Y_{t+k} - \mu)(Y_t - \mu)]$$

och för en stationär tidsserie gäller att

$$\text{Cov}(Y_{t_1+k}, Y_{t_1}) = \text{Cov}(Y_{t_2+k}, Y_{t_2})$$

ty $E(Y_{t+k}) = E(Y_t) = \mu$ enligt definitionen för stationaritet. Med andra ord är autokovariansen enbart en funktion av tidsförskjutningen k . Vidare gäller att autokorrelationen definieras

$$\rho_k = \frac{\gamma_k}{\sqrt{\gamma_0}\sqrt{\gamma_0}} = \frac{\gamma_k}{\gamma_0}$$

där γ_k är autokovariansen (som enbart är beroende av tidsförskjutningen k) och γ_0 är variansen (alltså autokovariansen utan tidsförskjutning) (ibid.).

2.5.1 Modelltester- och informationskriterier

För att jämföra och mäta anpassningen för sin modell mot sin data är Akaikes informationskriterium, AIC, ett vanligt tillvägagångssätt (Chan och Cryer, 2008). AIC väljer, med hjälp av log-likelihoodfunktionen, den modell av flera som har bäst anpassning till den underliggande datamaterialet. AIC skrivs på följande vis:

$$\text{AIC} : -2\log(\text{maximumlikelihood}) + 2k$$

Genom att jämföra värdet för AIC mellan olika modeller kan den bästa anpassningen hittas, ett lägre värde indikerar en bättre anpassad modell (ibid.). Termen k , vilken står för hur många parametrar som skattats i modellen och består av antingen $p + q$ eller $p + q + 1$, har till syfte att straffa modeller med fler parametrar än nödvändigt. Detta eftersom värdet på AIC stiger ju fler parametrar som finns (ibid.).

AIC har uppvisat brister för datamaterial med färre observationer genom att rekommendera modeller med för många parametrar, varför en korrigerad AIC, AIC_c, har föreslagits och i allt högre grad ersatt AIC (Chatfield, 2011). AIC_c består, i likhet med AIC, av en log-likelihoodfunktion och en straffterm k men inkluderar även antalet observationer n och skrivs:

$$\text{AIC}_c = \text{AIC} + \frac{2(k+1)(k+2)}{n-k-2}$$

I praktiken innebär detta att AIC_c straffar ett större antal parametrar hårdare än vad AIC straffar samma antal parametrar (ibid.).

2.6 Box och Jenkins metod

Box och Jenkins metod delas ibland in i 3 delar som består av

1. *Modellidentifiering – formulera en lämplig modell*
2. *Modellestimering - anpassa modellen*
3. *Modellkontroll – testa och, om det behövs, justera modellen*

Metoden brukar involvera att undersökaren initialt observerar sin tidsseriediagram för att avgöra om den finns trend- och säsongsmönster för att sedan differentiera eller säsongsdifferentiera denna tills stationäritet uppstår (Chatfield, 2011). Sedan brukar analyser av det stationära materialets autokorrelations- och partiella autocorrelationsplottar göras för att avgöra ordningen på parametrarna i ARIMA eller SARIMA-modellen. För att avgöra lämpligheten på den senare färdiga modellen görs kontroller av residualerna, det vill säga tester på de termer som inte förklaras av modellen (ibid.). I denna uppsats kommer en residualplott, ett histogram över residualerna samt en QQ-diagram för residualerna analyseras och kontrolleras mot normalitetsantagandet. För att se hur väl modellen passar underliggande data kommer informationsmättet AIC_c att användas.

En annan metod som kommer användas är ett så kallat test set, vilket innebär att en del av observationerna i tidsserien plockas bort under modellskapandet för att sedan läggas till när prognosen är gjord. Test set'et används då för att ge en indikation på hur väl modellen passar det verkliga datamaterialet. I denna uppsats plockas 10% bort vilket motsvarar 24 observationer (motsvarande 24 månader) (ibid.).

2.7 Prediktion

När modellskapandet anses färdigt kan prediktioner börja göras, i denna uppsats kommer precisa prediktionsvärden liksom prediktionsintervall om 99 respektive 95 procent göras (ibid.). Ett genomgående tema bland många prediktionsintervall för ARIMA-skapande är att de blir för snäva. Detta problem beskrivs, av bland annat av Chatfield, som ett resultat av att den underliggande modellen för datamaterialet förändras i takt med tiden (ibid.).

För denna uppsats kommer 4 år (48 månader) att predikteras. Det är viktigt att ha i åtanke att ju längre prediktioner som görs desto mer osäkra blir de, detta synliggörs bland annat genom prediktionsintervall som blir bredare och bredare ju längre fram i tiden som modellerna predikterar. För att kontrollera hur väl anpassad den skapade modellen är till materialet så tas 10%, 24 månader, av observationerna bort.

För att tydligare illustrera formlerna för prediktionen används en ARIMA(1,1,1) som exempel, vilken kan ses nedan

$$\text{ARIMA}(1,1,1) : Y_t = (1 + \phi_1)Y_{t-1} - \phi_1 Y_{t-2} + e_t - \theta_1 e_{t-1}$$

En prediktion av en ARIMA(1,1,1) modell tar följande uttryck, där $\hat{y}_N(h)$ representerar prediktionen h steg framåt (i vår analys $h = 48$) och N står för tiden (Chatfield, 2004).

$$\hat{y}_N(h) = \begin{cases} (1 + \phi_1)y_N - \phi_1 y_{N-1} - \theta_1 e_N & h = 1 \\ (1 + \phi_1)\hat{y}_N(1) - \phi_1 y_N & h = 2 \\ (1 + \phi_1)\hat{y}_N(h-1) - \phi_1 \hat{y}_N(h-2) & h \geq 3 \end{cases}$$

Detta leder till prediktionsintervallen för h steg framåt blir

$$\hat{y}_N(h) \pm z_{\alpha/2} \sqrt{\text{Var}[e_N(h)]}$$

där $z_{\alpha/2}$ antingen är 1.96 eller 2.58 beroende på om prediktionsintervallet är inom 95% eller 99%. Här är $e_N(h)$ prediktionsfelet gjort vid tid N när man predikterar h steg framåt (ibid.).

3. Data

Materialet som analyseras är nedladdat från Myndigheten för samhällsskydd och beredskaps statistikdataportal i form av månadsdata, innan nedladdning begränsades det för avsiktlig brand och var specificerat till enbart personbilar. Statistiken från MSB grundar sig i insats- och händelserapporter från räddningstjänsten. Sveriges kommuner och landsting (SKL) har en generell kommungruppsindelning där de tre huvudgrupperna kommer användas i denna uppsats, dessa presenteras nedan:

1. Storstäder och storstadsnära kommuner
Storstäder och Pendlingskommun nära storstad
2. Större städer och kommuner nära större stad
Större stad, Pendlingskommun nära större stad och Lågpendlingskommun nära större stad
3. Mindre städer/tätorter och landsbygdskommuner
Mindre stad/tätort, Pendlingskommun nära mindre stad/tätort, Landsbygdskommun, Landsbygdskommun med besöksnäring

(Sveriges kommuner och landsting, 2016).

Grupp 1 är städer med mer än 200 000 invånare samt städer med 40% utpendling till storstad, såsom Malmö och Staffanstorps. Grupp 2 innefattar städer med mellan 40 000 och 200 000 invånare och städer med utpendlare till större städer, såsom Lund och Sjöbo. I grupp 3 ingår exempelvis Karlshamn och Sölvesborg, alltså städer med mellan 15 000 och 40 000 invånare och landsbygdsorter med färre än 15 000 invånare (ibid.).

För att möjliggöra en tydligare jämförelse mellan områdena hämtas även datamaterialet ned som antal anlagda personbilsbränder per 1000 invånare. Denna data finns att hämta direkt i MSBs statistikdataportal, där är datan även ordnad enligt SKLs kommungruppsindelning. Exempelvis har datamaterialet för storstäder genererats genom att summera antalet bränder som skett i alla storstäder och pendlingskommuner nära storstad per månad för att sedan divideras med antalet skrivna där. Denna kvot har sedan multiplicerats med 1000. Denna beräkning har gjorts varje månad i samtliga huvudgrupper. Från detta datamaterial har det sedan skapats tre tidsserier, samtliga med 240 månadsobservationer från januari 1998 till december 2017, där 216 observationer användes under modellskapandet.

4. Beräkningar och resultat

I kommande avsnitt presenteras processen för modellbyggnad med hjälp av ACF (autokorrelationsplott) och PACF (partiell autokorrelationsplott). Även histogram och QQ-diagram kommer tas till hjälp för att avgöra ifall residualerna för modellerna följer normalfördelningsantagandet. Vid jämförelsen mellan de olika samhällsområdena kommer relativ data att användas, alltså antalet anlagda personbilsbränder beräknat per 1000 invånare för respektive samhällsområde.

Värt att poängtera är att inga outliers, det vill säga extraordinära mönster, kommer plockas bort eller manipuleras för att underlätta analysen. I situationer där outliers plockas bort beror det generellt på mänskliga misstag eller att insamlingen har felbehandlats, men då outliers i denna situationen bedöms vara resultat av yttre faktorer som hastigt ändrat människors beteende (och därmed orsakat fler bilbränder) behålls de.

Redan innan analysen påbörjades observerades att andelen bränder var överrepresenterade i storstäder. I storstäder sker hälften (50.2%) av bränderna samtidigt som bara 36% av invånarna bor där. I mindre städer var situationen omvänd, 25% bor där medan 13.3% av bränderna sker där. I större städer skedde ungefär lika många bränder som andel invånare som bor där (drygt 38% bor där och 36.5% av bränderna sker där).

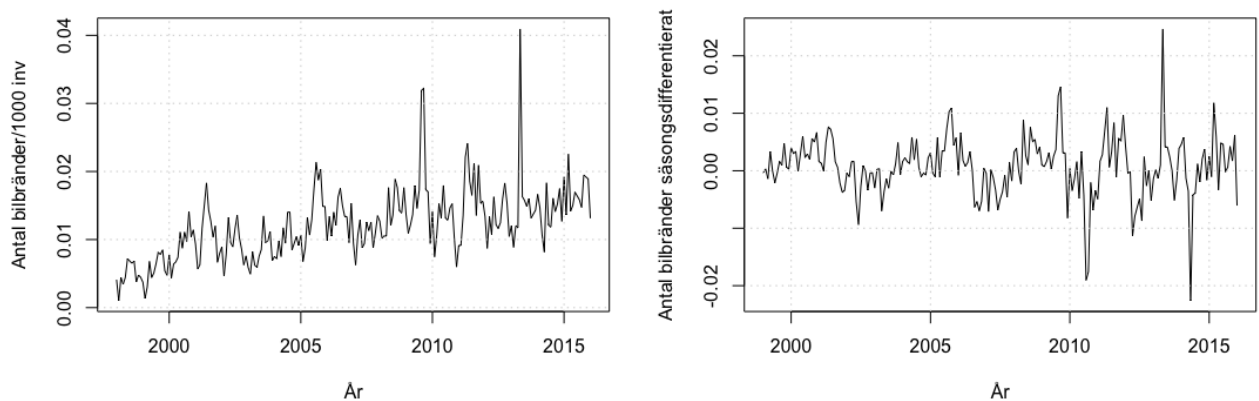
4.1 Storstäder och storstadsnära kommuner

4.1.1 Bakgrundsinformation

Tabell 4.1.1: Bakgrundsinformation

	Antal bilbränder	Antal bilbränder per 1000 inv
Medelvärde	39.3	0.0120
Standardavvikelse	18.5	0.0052
Min	3	0.0010
Max	143	0.0409

I denna grupp data ingår 3 städer (Stockholm, Göteborg och Malmö) samt 43 kommuner. Det totala antalet bränder var 8523 stycken mellan januari 1998 och januari 2016. Medelvärdet är 39.3 stycken anlagda bilbränder per månad och standardavvikelsen är 18.5. Detta ger en indikation på hur pass stor spridningen är i förhållande till medelvärdet. Siffrorna är beräknade av data fram till och med januari 2016, alltså exkluderat för test setet, genom att summera det totala antalet bränder för hela områdesgruppen och sedan dividera med antalet månader. Vidare har det i februari 1998 enbart registrerats 3 anlagda bilbränder. Maxvärdet 143 noterades i maj 2013. Medelvärdena för de olika månaderna varierar mellan 25.4 (0.0077) bränder per månad för februari och 48.8 (0.0151) för augusti, där 0.0077 och 0.0151 är medelvärdena för antalet bränder per 1000 invånare. Även dessa siffror är beräknade genom att summera antalet bränder för respektive månad och sedan dividera med de antalet månader som denna summa bygger på.



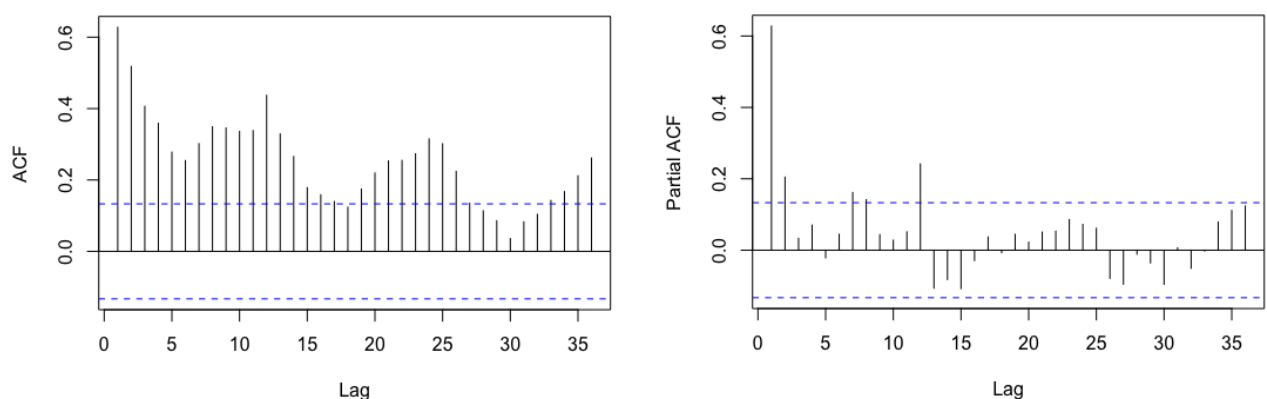
Figur 4.1.1: Tidsseriediagram över antalet relativa bilbränder i storstäder före (vänster) och efter (höger) säsongsdifferentiering

Som man kan observera i figur 4.1.1 är tidsintervallets maxvärde, det vill säga 0.0409 bilbränder per 1000 invånare i maj 2013 (se Max i tabell 4.1.1), en tydlig outlier. Orsaken är bland annat upplopp i Stockholm natten mot den 23e maj som orsakade flertalet uttryckningar. Som jämförelse kan nämnas att i maj 2012 och maj 2014 drabbades Stockholm av 26 respektive 29 bilbränder, den siffran var i maj 2013 80 stycken bilbränder.

I figur 4.1.1 kan man, precis som man kan observera outliers, även se en uppåtgående trend och att tidsserien inte är stationär, vilket också bekräftas av Dickey-Fullers test som blev $t_{obs} = -3.1788$ med p-värde 0.09263. Genom att säsongsdifferentiera tidsserien uppnåddes ett resultat på $t_{obs} = -5.4155$ med ett p-värde mindre än 0.01, vilket gjorde att processen att anpassa en SARIMA-modell kunde fortsätta.

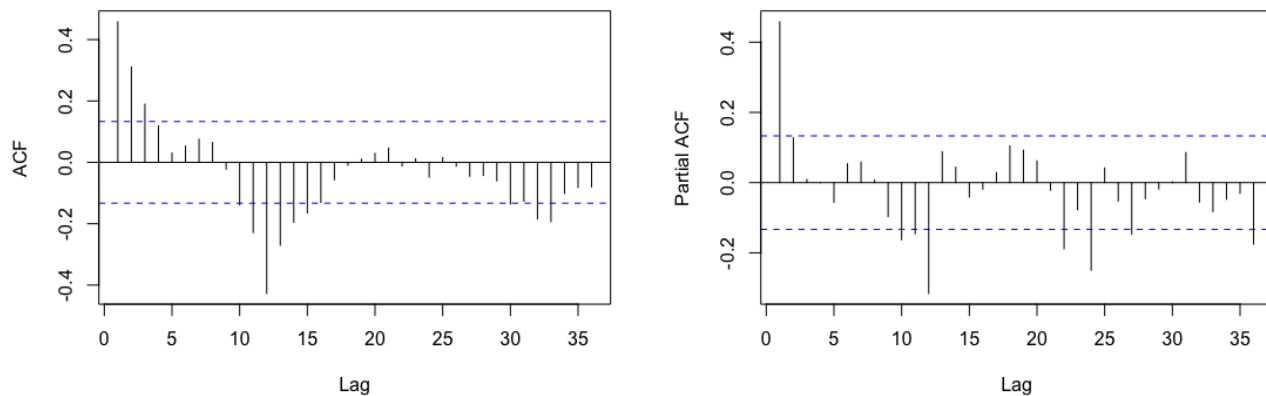
4.1.2 Modellskapande

ACF och PACF



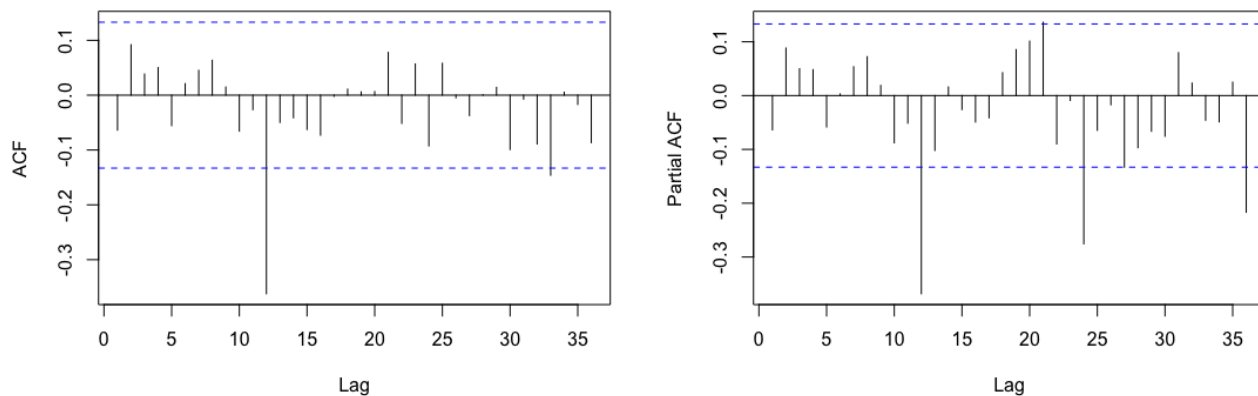
Figur 4.1.2: ACF och PACF över residualerna före säsongsdifferentiering

Som man kan se verkar det finnas ett säsongsberoende som yttrar sig i spikar var 12e lagg, vilket, i kombination med tidsseriediagrammet i figur 4.1.1, motiverar en säsongsdifferentiering för att kunna fortsätta skapa SARIMA-modellen.



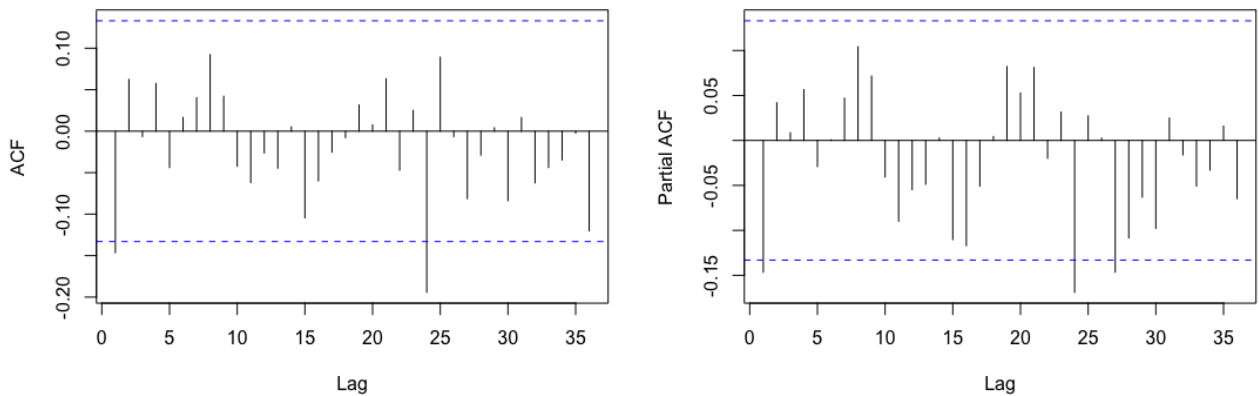
Figur 4.1.3: ACF och PACF över residualerna efter säsongsdifferentiering

Eftersom ACF:en avtar exponentiellt i kombination med att PACF:en har en spik i 1:a lagget så tillsätts en AR(1)-process.



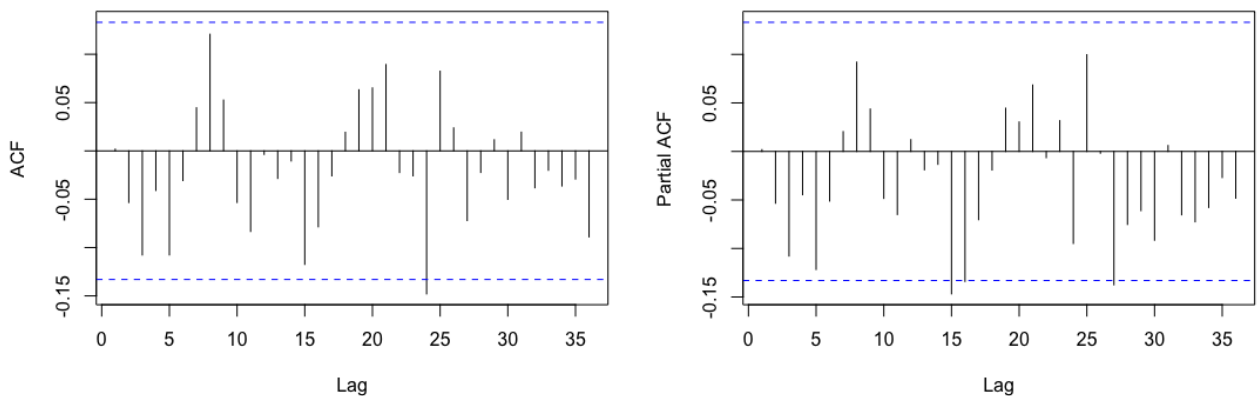
Figur 4.1.4: ACF och PACF över residualerna för en SARIMA(1,0,0)(0,1,0)₁₂

ACF:en uppvisar en spik i 12:e tidslagget medan PACF:en har ett avtagande mönster var 12e tidsförskjutning, man kan tydligt avläsa detta i lagg 12, 24 och 36. Detta är tecken på en SMA(1)-process varför en SARIMA(1,0,0)(0,1,1)₁₂ också prövades.



Figur 4.1.5: ACF och PACF över residualerna för en $SARIMA(1,0,0)(0,1,1)_{12}$

Eftersom man kan se en spik i första tidsförskjutningen för ACF:en så tillsattes en MA(1)-process.



Figur 4.1.6: ACF och PACF över residualerna för en $SARIMA(1,0,1)(0,1,1)_{12}$

Här bedöms modellen vara tillräckligt slumpmässig eftersom det inte går att urskilja några särskilda mönster. Under processens gång har följande AIC_c observerats.

Tabell 4.1.2: AIC_c värden

Modell	AIC_c
$SARIMA(0, 0, 0)(0, 1, 0)_{12}$	-1569.52
$SARIMA(1, 0, 0)(0, 1, 0)_{12}$	-1618.24
$SARIMA(1, 0, 0)(0, 1, 1)_{12}$	-1679.04
$SARIMA(1, 0, 1)(0, 1, 1)_{12}$	-1685.62

AIC_c har minskat från -1569.52 till -1685.62 vilket tyder på att modellen blivit bättre anpassad till datamaterialet.

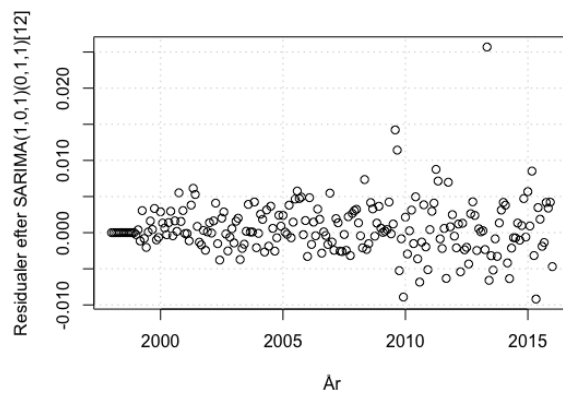
Modellparametrar

Som framkommer av tabellen ovan är samtliga av modellens koefficienter signifikanta. Ytterligare ett sätt att kontrollera trovärdigheten i modellen är genom att undersöka ifall residualerna ser slumpmässiga och normalfördelade ut.

Tabell 4.1.3: Modellparametrar

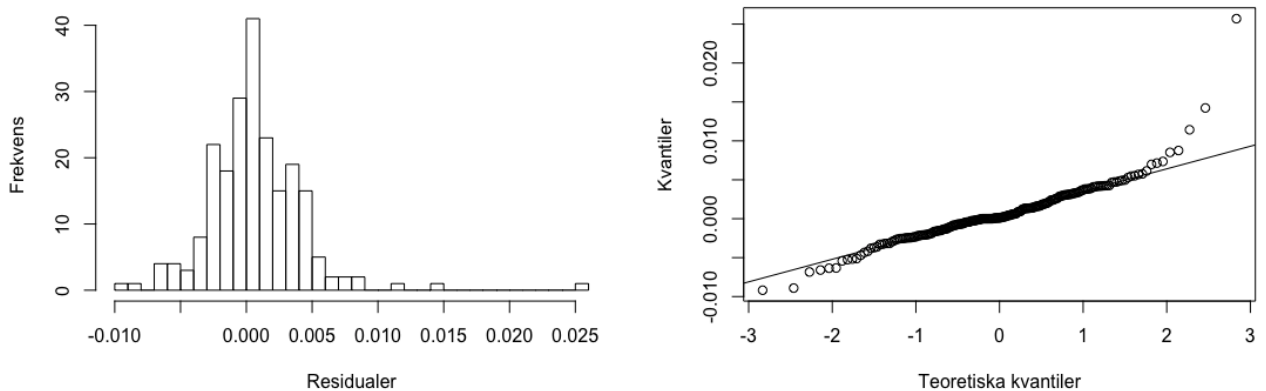
Modell	Koefficienter	P-värde	Standardavvikelse
ϕ_{AR}	0.841663	$< 2.2 \cdot 10^{-16}$ ***	0.097651
θ_{MA}	-0.48551	0.002487 **	0.158861
θ_{SMA}	0.765476	$< 2.2 \cdot 10^{-16}$ ***	0.067770

Residualtest



Figur 4.1.7: Modellens residualer uppritade över tid

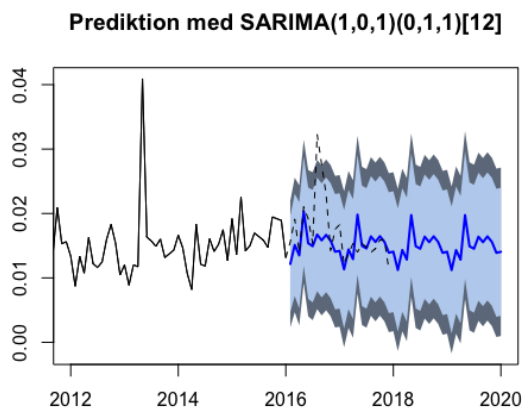
Residualerna uppritade över tid visar på en svagt ökande spridning, sannolikt beroende på några få outliers.



Figur 4.1.8: Histogram och QQ-diagram över modellens residualer

Histogrammet är relativt normalfördelat med undantag för några outliers som orsakar en skevhet i fördelningen, samma problem orsakar även skevheten och de avvikande observationerna i QQ-diagrammen, modellen accepteras trots detta och prognostiseringen kan ses nedan.

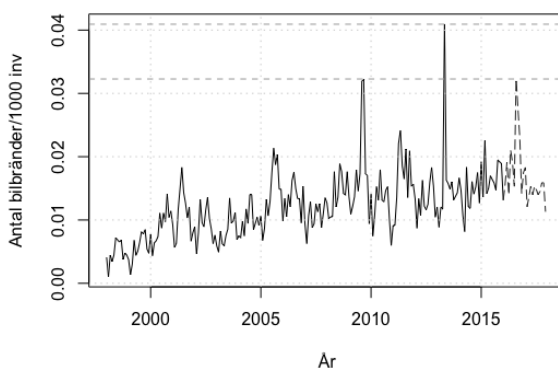
4.1.3 Prediktion



Figur 4.1.9: Prognos för 4 år framåt

Prognostiseringen gjordes för 4 år från och med februari 2016 till och med januari 2020. Den streckade linjen symboliserar de observationer som inledningsvis plockades bort, test setet, och kan alltså ses som en typ av referens för hur väl modellen förutspår de första två åren. Det ljusgrå området representerar konfidensintervallet om 99%, det mörkgrå konfidensintervallet om 95% medan den blå linjen är prediktionen. För att tydliggöra prediktionen har enbart åren 2012 till 2020 tagits med i plotten, samtliga år (1998-2015) har dock använts som underlag till SARIMA-skapandet.

Som kan noteras täcker varken det 95%-iga eller 99%-iga prediktionsintervallet in de verkliga värdena för två observationer (se rödmarkerade värden i bilaga 1), medan resterande värden rätt så väl följer prediktionen. Att modellen missar två observationer kan indikera en modell som inte fungerar särskilt väl. Toppnoteringen 0.0323 för de streckade området, eller 118st bränder, under Augusti 2016 är det näst högsta värdet som noterats under MSBs datainsamling för Storstäder. Nedanstående figur 4.1.10 förtydligar detta.



Figur 4.1.10: Tidsseriediagram för att tydliggöra skillnaden i maximumvärden

Den högre streckade linjen tangerar det högsta värdet för perioden som inte ingick i modellskapandet och representerar också hela periodens näst högsta värde. De tre toppnoteringarna i september 2009 (106st), maj 2013 (143st, varav 80 st i Stockholm) och augusti 2016 (118st) är samtliga resultat av oroligheter på olika platser i landet som orsakat långt över förväntat antal bilbränder. I september (och augusti) 2009 var det oroligheter i Göteborg och i augusti 2016 skedde liknande händelser i Malmö.

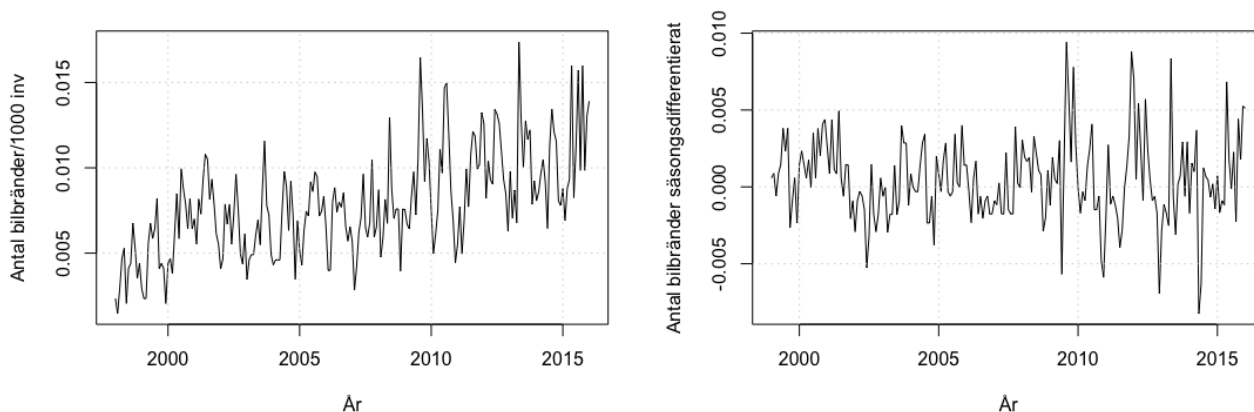
4.2 Större städer och kommuner nära större stad

4.2.1 Bakgrundsinformation

Tabell 4.2.1: Bakgrundsinformation

	Antal bilbränder	Antal bilbränder per 1000 inv
Medelvärde	28.0	0.0078
Standardavvikelse	11.3	0.0030
Min	5	0.0015
Max	64	0.0174

Större städer och kommuner nära större stad inkluderar 21 större städer samt 87 kommuner med pendling till större städer. Det noterades 6068 bränder för hela området mellan 1998 och 2016. För denna gruppen var variationen kring medelvärdet, det vill säga standardavvikelsen, något mindre än för storstäder. Detta beror sannolikt på att antalet anlagda bilbränder generellt är lägre här. Som mest observerades 64 bilbränder under en månad, maj 2013, minimumvärde var 5 bilbränder vilket inträffade i februari 1998. Medelvärdena varierade mellan 18.7 (0.0052) bränder för februari och 36.2 (0.0101) för augusti.

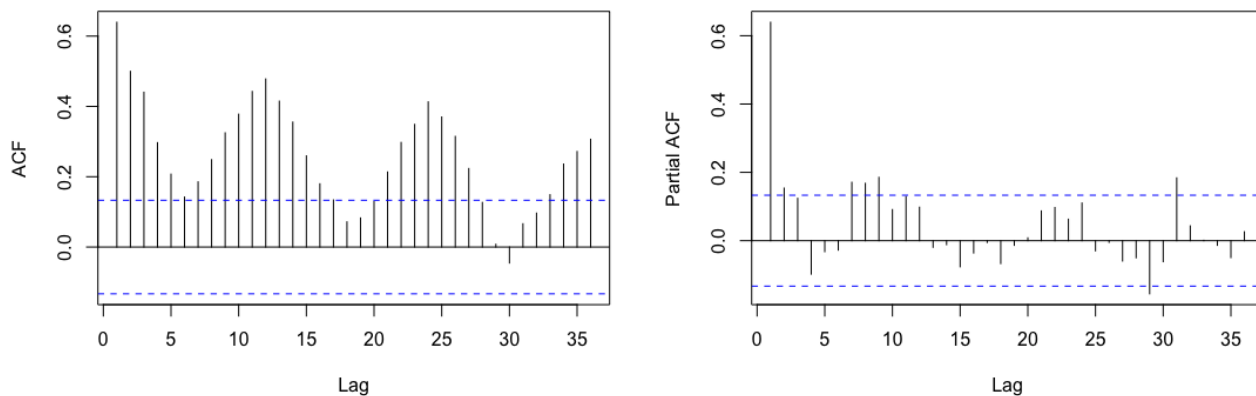


Figur 4.2.1: Tidsseriediagram över antalet relativa bilbränder i storstäderna före (vänster) och efter (höger) säsongsdifferentiering

Ovanstående tidsseriediagram, figur 4.2.1, visar en positivt ökande men varierande trend. Det finns relativt stora variationer kring en svag uppåtgående trend vilket indikerar att serien inte är stationär. Dickey-Fullers test bekräftar att tidsserien inte är stationär med $t_{obs} = -2.8814$ med p-värde 0.2062. Efter en säsongsdifferentiering blev $t_{obs} = -5.1877$ med ett p-värde på mindre än 0.01 och tidsserien kan alltså anses vara stationär.

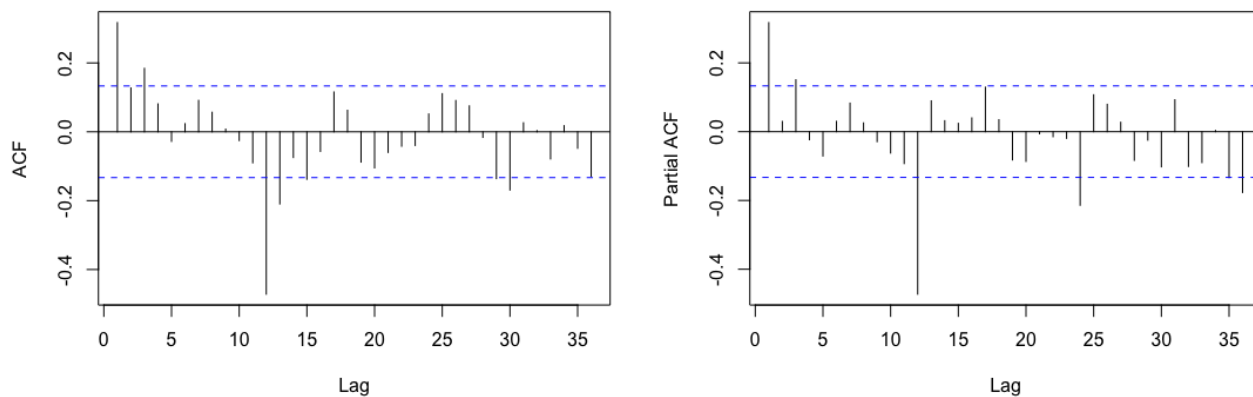
4.2.2 Modellskapande

ACF och PACF



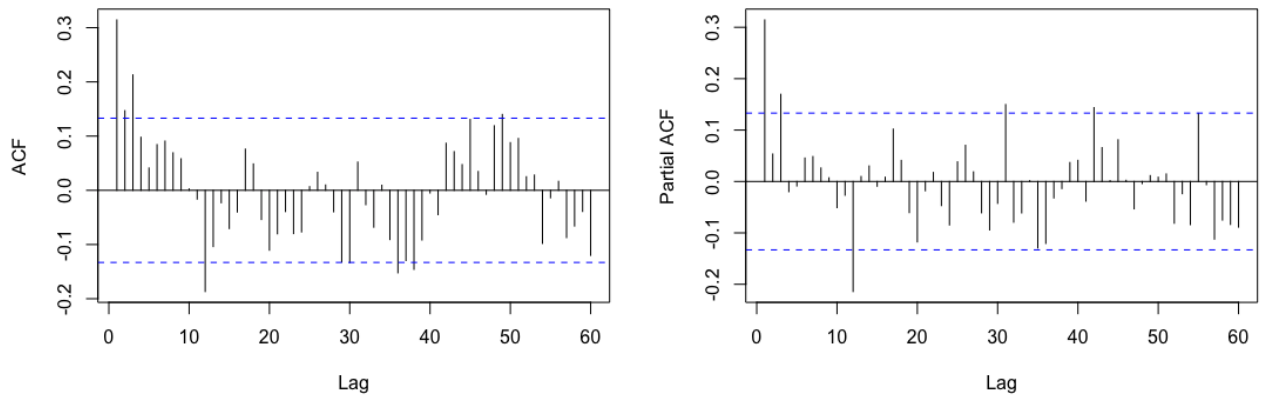
Figur 4.2.2: ACF och PACF över residualerna före säsongsdifferentiering

Även i ACF:en i figur 4.2.2 kan ett säsongberoende observeras vilket återigen motiverar en säsongsdifferentiering för att uppnå stationäritet.



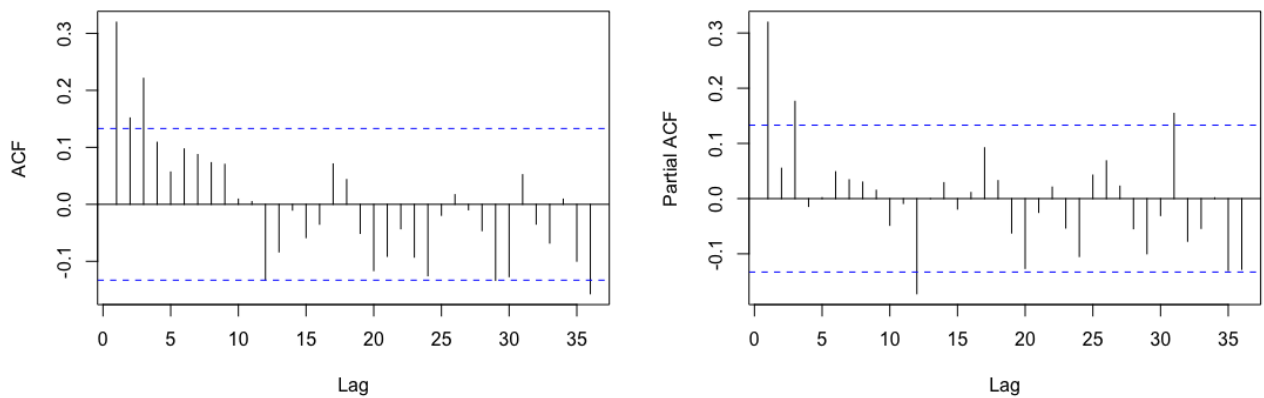
Figur 4.2.3: ACF och PACF över residualerna efter säsongsdifferentiering

På ACF:en kan man observera spik på lag 12 och 24, vilket, tillsammans med PACF:ens avtagande mönster längs med vart 12:e tidsförskjutning indikerar en SMA(1)-process.



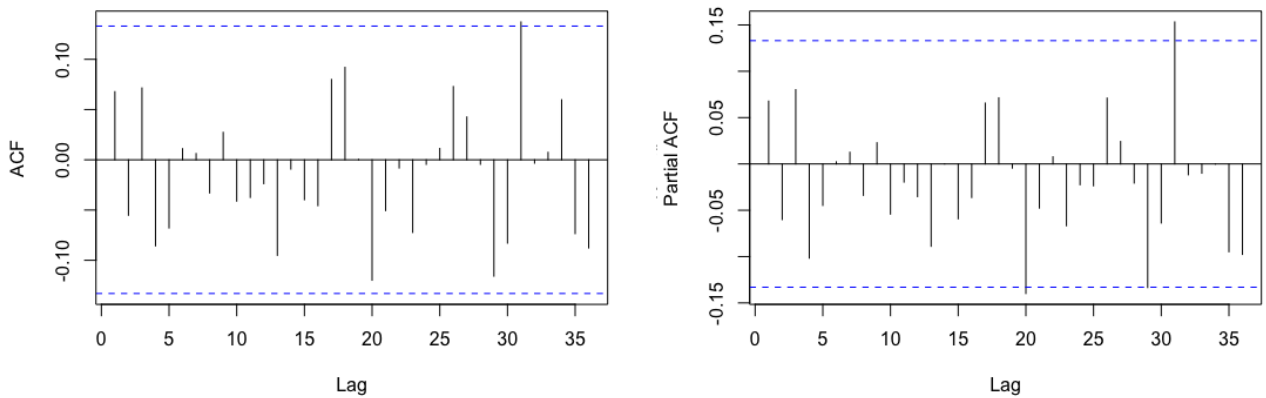
Figur 4.2.4: ACF och PACF över residualerna för en $SARIMA(0,0,0)(0,1,1)_{12}$

ACF:en är exponentiellt avtagande var 12e tidsförskjutning vilket tyder på en SAR(1)-process.



Figur 4.2.5: ACF och PACF över residualerna för en $SARIMA(0,0,0)(1,1,1)_{12}$

Eftersom både ACF och PACF här har en spik i första lagg, samt att de båda avtar någorlunda över tid så testas en ARMA(1,1) att läggas till.



Figur 4.2.6: ACF och PACF över residualerna för en $SARIMA(1,0,1)(1,1,1)_{12}$

Residualerna bedöms här vara tillräckligt slumpmässiga eftersom det inte går att urskilja några mönster. Följande AIC_c har observerats.

Tabell 4.2.2: AIC_c värden

Modell	AIC_c
$SARIMA(0,0,0)(0,1,0)_{12}$	-1832.44
$SARIMA(0,0,0)(0,1,1)_{12}$	-1889.26
$SARIMA(0,0,0)(1,1,1)_{12}$	-1888.23
$SARIMA(1,0,1)(1,1,1)_{12}$	-1945.46

Att AIC_c minskat från -1832.44 till -1945.46 indikerar att modellen blivit bättre anpassad till datamaterialet.

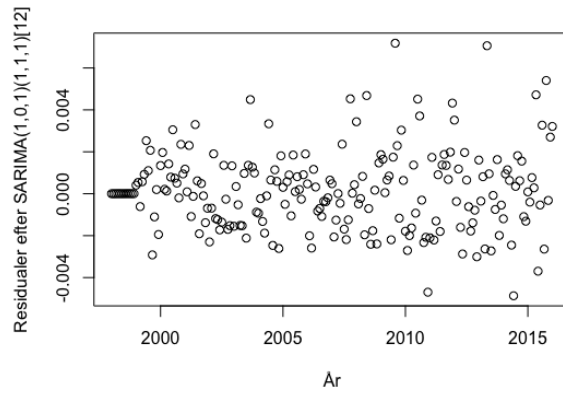
Modellparametrar

Tabell 4.2.3: Modellparametrar

Modell	Koefficienter	P-värde	Standardavvikelse
ϕ_{AR}	0.995886	$< 2.2 \cdot 10^{-16}$ ***	0.011234
θ_{MA}	-0.796758	$< 2.2 \cdot 10^{-16}$ ***	0.059386
ϕ_{SAR}	-0.071152	0.4098	0.086331
θ_{SMA}	-0.896730	$< 2.2 \cdot 10^{-16}$ ***	0.106417

Trots att samtliga koefficienter inte är signifikanta anser vi att modellen är godkänd. Nästa steg består i att undersöka residualerna.

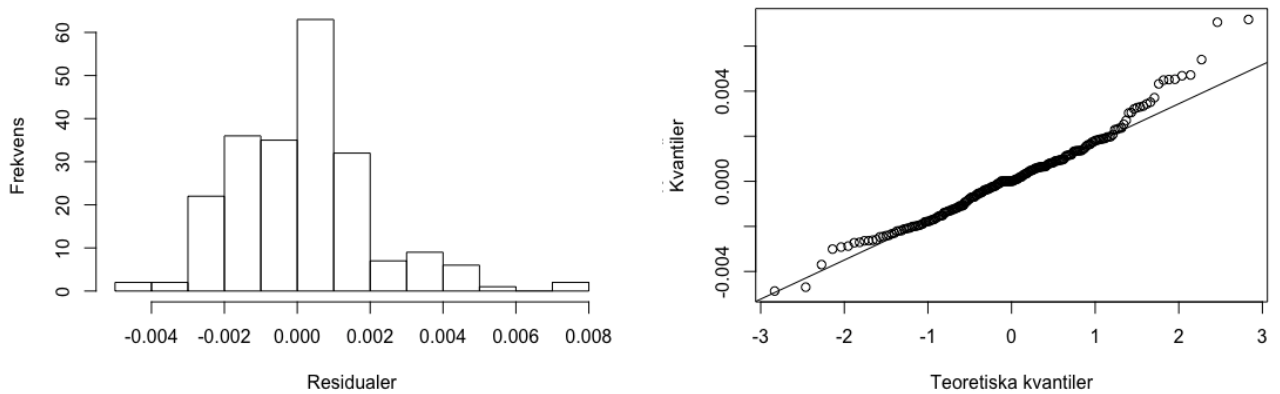
Residualtest



Figur 4.2.7: Modellens residualer uppritade över tid

En plott av residualerna uppritat mot tiden uppvisar inga tydliga mönster, histogram och QQ-diagram analyseras för att få en tydligare bild.

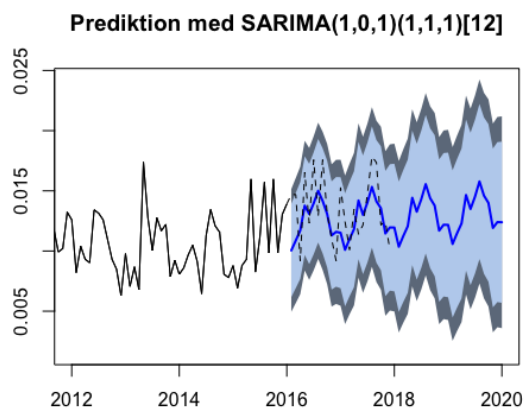
Residualtest



Figur 4.2.8: Histogram och QQ-diagram över modellens residualer

Modellens residualer ser inte ut att följa ett normalfördelningsantagande, det finns en skevhet i både histogrammet och QQ-diagrammet.

4.2.3 Prediktion



Figur 4.2.9: Prognos för 4 år framåt

Även här gjordes prognosen från och med feburari 2016 till och med januari 2020. Precis som tidigare är den streckade svarta linjen test setet. Det ljusgråa området representerar konfidensintervallet om 99%, medan det ljusgrå är det 95%-iga konfidensintervallet. Den blå linjen är prediktionen. Även här har enbart åren 2012 till 2020 tagits med i plotten för att tydliggöra prediktionen, samtliga år (1998–2015) har dock använts som underlag i modellskapandet.

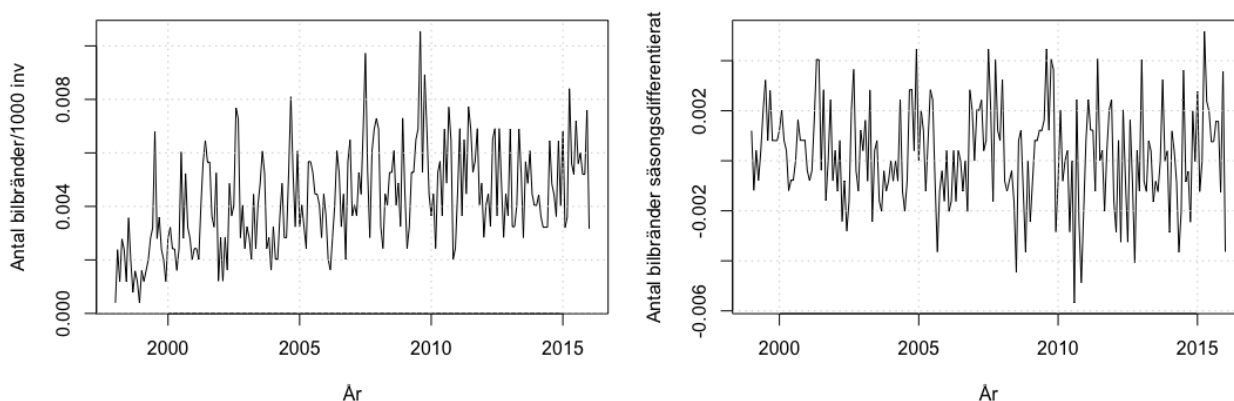
4.3 Mindre städer och landsbygd

4.3.1 Bakgrundsinformation

Tabell 4.3.1: Bakgrundsinformation

	Antal bilbränder	Antal bilbränder per 1000 inv
Medelvärde	10.5	0.0043
Standardavvikelse	4.6	0.0019
Min	1	0.0004
Max	26	0.0105

I likhet med för föregående områden påbörjades analysen med att få en överblick av materialets deskriptiva statistik. Denna grupp inkluderar 29 mindre städer, 52 pendlingskommuner samt 55 landsbygdskommuner. Som väntat, ty mindre städer och landsbygden har färre invånare, är förekomsten av anlagda bilbränder här lägre än för både större städer och storstäder. Tydligast bevis för detta fås genom att jämföra medelvärdena per månad, som med 10.5 bränder per år för mindre städer och landsbygder är mindre än för både större städer och storstäder där de i snitt varit 28.0 respektive 39.3 anlagda bilbränder per månad. Även de totala antalet bränder, 2262, vittnar om att det sker avsevärt färre antal bränder i mindre städer och landsbygd än i storstäder och större städer. För mindre städer varierade medelvärdena månaderna emellan mellan 7.9 (0.0027) för mars och 13.8 (0.0056) för juli. Innan modellskapandet kan starta tas återigen 10%, 24 observationer, bort för att senare kunna se avgöra hur väl modellens prediktion följer de verkliga värdena.



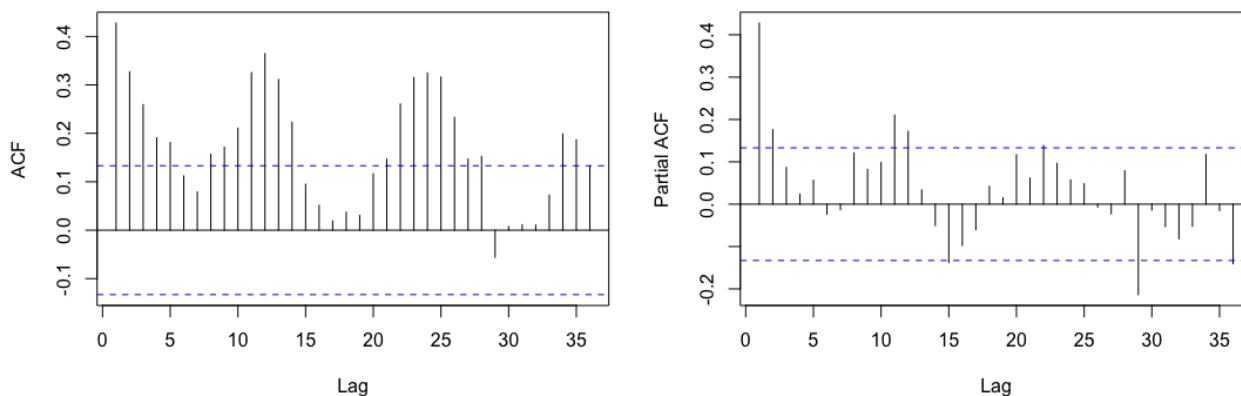
Figur 4.3.1: Tidsseriediagram över antalet relativa bilbränder i mindre städer före (vänster) och efter (höger) säsongsdifferentiering

Figur 4.3.1 för det relativa antalet bilbränder uppvisar en svag uppåtgående trend och är inte stationär i dess ursprungliga form, varför en säsongsdifferentiering utfördes. Säsongsdifferentieringen visade sig ge tidsserien stationäritet. Till skillnad från storstäder och större städer är den ursprungliga tidsserien för mindre städer och landsbygd mer varierande och mindre förutsägbar.

Dickey-Fullers test av stationäritet bekräftade att den ordinarie tidsserien inte var helt stationär och innan differentiering uppmättes ett t_{obs} på -2.9349 med p-värdet 0.1838 . Efter säsongsdifferentiering uppmättes $t_{obs} = -6.686$, med ett p-värde om mindre än 0.01 , vilket alltså är lägre än den kritiska gränsen på -1.96 som gäller på 5%-nivån. Efter att antagandet för stationäritet uppfyllts kan, som nämnt, processen för bildandet av en modell starta.

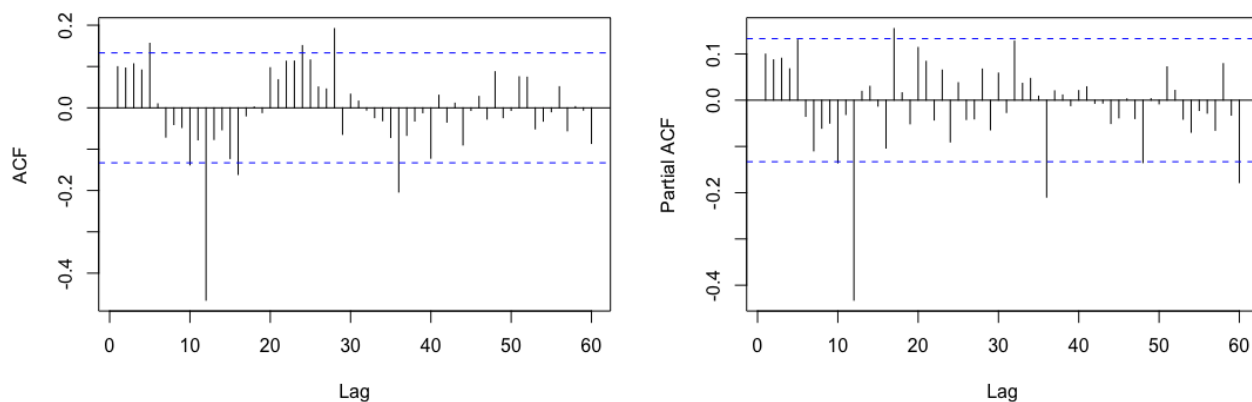
4.3.2 Modellskapande

ACF och PACF



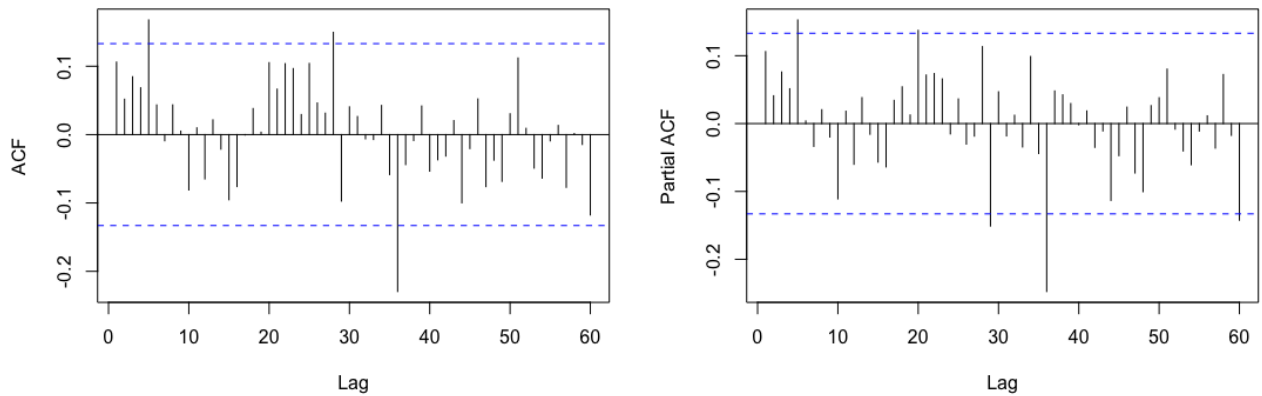
Figur 4.3.2: ACF och PACF över residualerna före säsongsdifferentiering

Initialt ritades ACF och PACF upp för att se om det förelåg några tydliga mönster. Eftersom ACF:en påvisar spikar vid lagg 12, 24 och 36 indikeras det att även här var säsongsdifferentiering ett passande metod.



Figur 4.3.3: ACF och PACF över residualerna efter säsongsdifferentiering

Till skillnad från analyserna utav storstäder och större städer så visar varken ACF eller PACF någon spik i lagg 1, däremot kan flera spikar identifieras i PACF:en, en på respektive lagg 12, 36, 48 och 60 osv. Detta motiverar att applicera en SMA(1) till modellen.



Figur 4.3.4: ACF och PACF över residualerna för en $SARIMA(0, 0, 0)(0, 1, 1)_{12}$

Efter flertalet försök att förbättra modellen bedöms till sist denna version vara bäst lämpad för fortsatt modellerande, spikarna till trots. Ett tydligt mönster går inte att utläsa och processen fortsätter. Troligtvis beror svårigheten i att försöka ansätta en passande modell på att ursprungsvariationen för materialet är av en annan typ än den för storstäder och större städer. AIC_c har minskat från -1964.71 till -2019.75

Tabell 4.3.2: AIC_c värden

Modell	AIC_c
$SARIMA(0, 0, 0)(0, 1, 0)_{12}$	-1964.71
$SARIMA(0, 0, 0)(0, 1, 1)_{12}$	-2019.75

vilket tyder på en bättre lämpad modell.

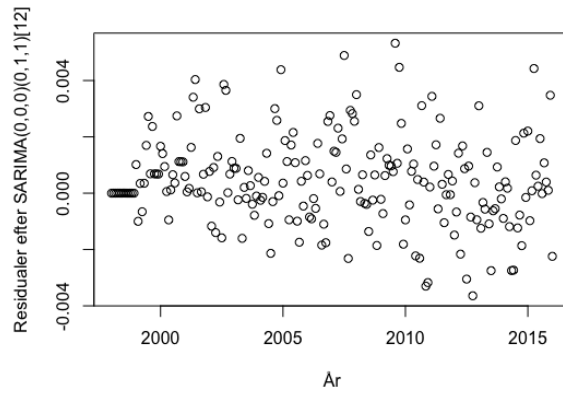
Modellparametrar

Tabell 4.3.3: Modellparametrar

Modell	Koefficienter	P-värde	Standardavvikelse
θ_{SMA}	-0.63661	$< 2.2 \cdot 10^{-16}$ ***	0.067109

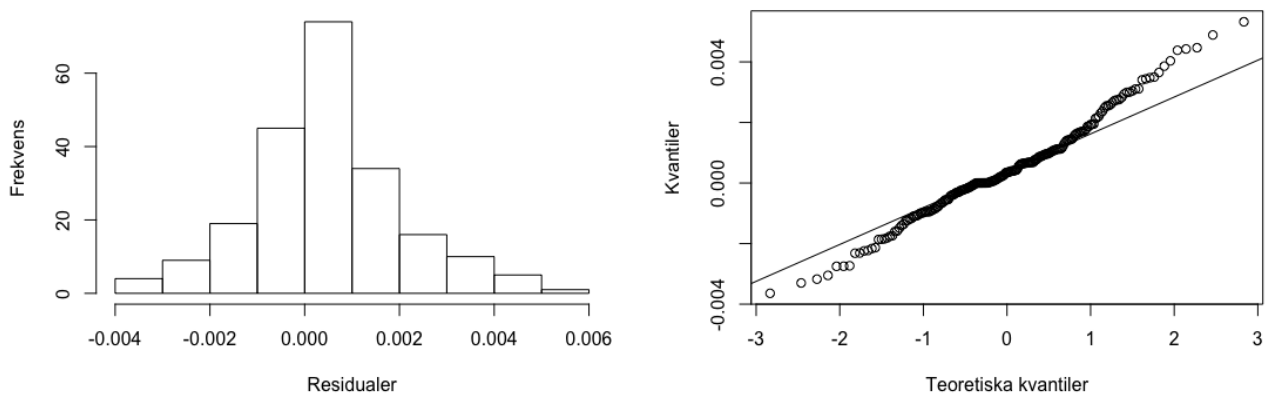
θ_{SMA} - koefficienten är signifikant på högsta signifikansnivå.

Residualtest



Figur 4.3.5: Modellens residualer uppritade över tid

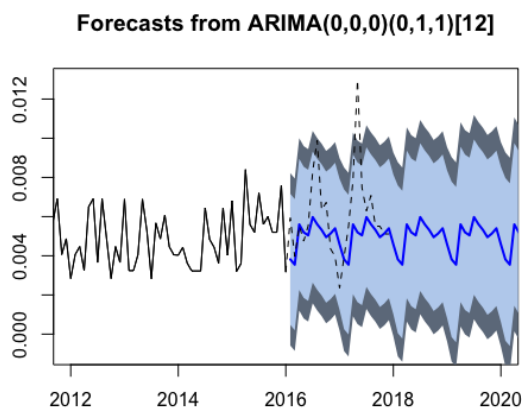
Residualerna ser slumpmässiga ut.



Figur 4.3.6: Histogram och QQ-diagram över modellens residualer

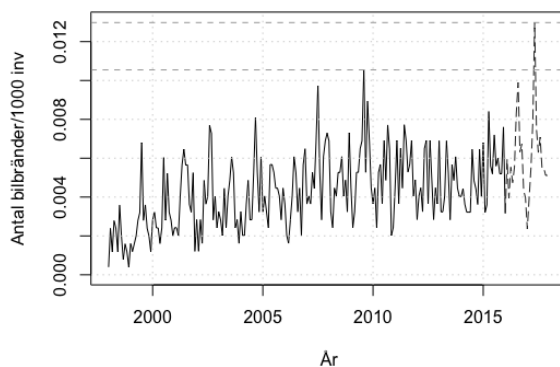
Fördelningen för residualerna från denna modell återges i histogrammet i figur 4.3.6. Motsvarande QQ-diagram visar på en viss skevhet i fördelningen.

4.3.3 Prediktion



Figur 4.3.7: Prognos för 4 år framåt

Prediktionen underskattar de verkliga värdet, även på 99%-nivån, vilket indikerar en dåligt anpassad modell. Dock skall poängteras att värdet som registrerades i maj 2017 (33 bränder, 0.0129 bränder per 1000 invånare), maxvärdet i det streckade området, är den enskilt högsta månadsnoteringen som gjorts för mindre tätorter sedan insamlandet av denna statistik startades, vilket sannolikt till stor del förklarar varför modellen misslyckats med att förutspå detta värde.



Figur 4.3.8: Tidsseriediagram för att tydliggöra skillnaden i maximumvärden

För att ytterligare tydliggöra hur pass stor denna maxnotering är, visas i figur 4.3.8 ett tidsseriediagram över hela datamaterialet. Den undre horisontala streckade linjen tangerar det högsta värdet innan 2017 (augusti 2009) medan den övre streckade linjen tangerar toppnoteringen i maj 2017. Som bekant beskriver standardavvikelsen (här 0.0019, se tabell 4.3.1) hur mycket observationerna avviker från medelvärdet (här 0.0043). Detta betyder att för denna tidsserie så förväntas observationerna variera mellan 0.0024 och 0.0062 bilbränder per 1000 invånare. Med detta i åtanke förstås hur osannolikt värdet 0.0129 är. Standardavvikelsen, 4.6, samt medelvärdet 10.5 indikerar, återigen, att denna högsta notering inträffat som ett resultat av ett statistiskt osannolikt scenario.

5. Diskussion och slutsatser

I detta arbete har tre olika prediktioner gjorts för att försöka förutspå antalet framtida anlagda bilbränder per 1000 invånare för tre olika områden i samhället. Under analysens gång har ett mönster som indikerar säsongstendenser såväl som uppåtgående trender observerats. Vidare har, för samtliga samhällsområden, ett högre antal anlagda bilbränder generellt gällt för sommarmånaderna maj, juni, juli och augusti. En tänkbar anledning till detta kan vara att fler människor rör sig utomhus på grund av varmare väder liksom att människor i högre utsträckning är lediga.

För både mindre städer och storstäder täckte prediktionsintervallet inte de borttagna verkliga observationerna, något som tyder på att verkligheten präglats av flera extrema noteringar under test setets gång. För mindre städer innehöll de verkliga borttagna värdena den enskilt högsta noteringen för hela tidsperioden, vilket även en mer avancerad modell av detta slag sannolikt inte hade kunnat förutspå (se figur 4.3.8). I bilaga 1 och 3 har de värden som inte det 99%-iga prediktionsintervallet täckt in rödmärkerats för att tydliggöra för läsaren att missarna i predktionen enbart beror på just dessa extremvärden.

Att förstå de bakomliggande orsakerna till sådana här förändringar är såklart av intresse. Eftersom *anlagda* bilbränder är studiens behandlingsämne, det vill säga bränder som människor startat *med syfte* att förstöra, är orsaken till bränderna i högsta möjliga mån beroende av just människors beteende. Som bekant är människors beteende ett resultat av en mängd faktorer, däribland mönster och beteende från människor runtomkring. En människa som väljer att anlägga en bilbrand kan därför komma att inspirera andra såväl som skapa en otrygg stämning i området denne lever i, vilken kan generera ytterligare bilbränder. Att brott av denna typen sällan klaras upp gör också att avsaknaden av rädsla för att gripas inte stoppar förövarna i särskilt hög utsträckning. Som tidigare nämnt skulle en annan tänkbar anledning kunna vara den ekonomiska vinningen man kan komma att få om man bränner upp sin egen bil i syfte att begå försäkringsbrott.

För storstäder var situationen liknande den för mindre städer, samtliga observationer som modellen missade att prediktera var resultat av yttre faktorer som i hög grad påverkat människors beteende. Det övergripande temat, att det under de senaste 5 åren observerats många extraordinära noteringar, säger möjligtvis något om ett samhälle i förändring. Att dessa noteringar gjorts för både mindre och stora städer indikerar också att problemet berör stora delar av samhället och således alltså är en angelägenhet för många människor. Generellt gäller för samtliga områden att prediktionsintervallen är relativt breda och därför kan anses mindre relevanta i praktiken. Detta beror troligtvis på att variansskattningen, vilken påverkas av nämnda extremnoteringar, är stor. Att undvika denna typ av variansskattning är eventuellt en uppgift för framtida forskning, kanske skulle borttagande eller utjämning av outliers vara en väg att gå. Att ta logaritmen av antalet bilbränder per 1000 invånare skulle även det kunna fungera innan en modellering skulle påbörjas. När man inför en logaritmisk skala blir nämligen värdena inte lika extrema.

Framtida forskning skulle även kunna inkludera mer förklarande variabler i sin analys och på så vis få en klarare bild av vad som ligger bakom denna typen av brott. Exempelvis hade jämförelser av området med olika utbildningsnivå, genomsnittsålder och andel arbetslösa varit intressant att analysera. Även anledningen till varför residualerna för modellerna inte fyllde kriterierna för normalfördelningsantagandet är också det ett intressant ämne att undersöka närmare. Myndigheten för samhällsskydd och beredskap för även statistik på tidpunkter under dagen för insatser såväl som veckodagar, båda lämpliga kandidater för vidare studier och analyser i ämnet. Även vid vilken byggnad eller i vilket område branden startats, såsom vid ett flerbostadshus, bensinstation eller på öppen yta, registreras i deras databas vilket skulle kunna komma att utgöra en intressant grund för framtida arbeten. Generellt för denna typen av undersökningar gäller att ett större urval ger bättre prediktioner eftersom undersökaren har mer information att gå på. Alltså hade en liknande undersökning kunnat göras om några år och sannolikt renderat en bättre modell.

Avslutningsvis är det av intresse att poängtera att prediktionerna för samtliga områden pekade på en uppåtgående trend. Med andra ord finns det i materialet underlag för att det existerar en ständig ökning för denna typen av händelser. Med den bristande uppklaringsgraden och de höga samhällskostnaderna i åtanke blir det sannerligen intressant att följa den framtida utvecklingen av bilbränder i Sverige.

Litteraturförteckning

- Bisgaard, S. och K. Murat (2011). *Time Series Analysis*. Hoboken: John Wiley och Sons.
- Box, G., G. Jenkins och G.C. Reinsel (2013). *Time Series Analysis: Forecasting and Control*. Hoboken: John Wiley och Sons.
- Chan, K-S. och J.D. Cryer (2008). *Time Series Analysis*. 2a upplagan. New York: Springer.
- Chatfield, C. (2004). *The Analysis of Time Series*. Boca Raton: Chapman Hall/CRC.
- Chatfield, C. (2011). *Time Series Forecasting*. Boca Raton: Chapman Hall/CRC.
- Kudo, Per (2018). *Siffrorna Avslöjar: Bränna bilar är ett nästan riskfritt brott*. URL: <https://www.svd.se/statistik-branna-bilar-ar-ett-nastan-riskfritt-brott> (hämtad 2018-12-08).
- Lind, Katarina (2016). *Kraftig Ökning av Bilbränder - kan inspirera andra*. URL: <https://www.svd.se/kraftig-okning-av-anlagda-bilbrander> (hämtad 2019-02-26).
- Myndigheten för samhällsskydd och beredskap (2011). *Samhällets kostnader för olyckor*. URL: <https://www.msb.se/ribdata/filer/pdf/26114.pdf> (hämtad 2019-01-22).
- Myndigheten för samhällsskydd och beredskap (2014). *Samhällets kostnader för att förebygga bränder*. URL: <https://www.msb.se/RibData/Filer/pdf/27409.pdf> (hämtad 2019-01-22).
- Myndigheten för samhällsskydd och beredskap (2015). *Räddningstjänst i siffror 2014*. URL: <https://www.msb.se/RibData/Filer/pdf/27615.pdf> (hämtad 2019-01-22).
- Nationellt centrum för lärande från olyckor (2008). *Bränders samhällsekonomiska kostnader*. URL: <https://www.msb.se/RibData/Filer/pdf/24341.pdf> (hämtad 2019-01-22).
- Nygren, Sanna (2016). *Nytt Mönster för Bilbränderna i Malmö*. URL: <https://www.aftonbladet.se/nyheter/kaeOL9/nytt-monster-for-bilbrander-na-i-malmo> (hämtad 2019-02-26).
- Shumway, R.H. och D.S. Stoffer (2011). *Time Series Analysis and Its Applications: With R Examples*. 3e upplagan. New York, NY: Springer Science + Business Media, LLC.
- Svensk Försäkring (2017). *Försäkringsbedrägerier i Sverige 2017*. Stockholm: Svensk Försäkring Larmtjänst AB. URL: <https://www.svenskforsakring.se/globalassets/rapporter/forsakringsbedragier/forsakringsbedragier-i-sverige-2017.pdf> (hämtad 2019-01-15).
- Sveriges kommuner och landsting (2016). *Kommungruppsindelning 2017*. Stockholm: Sveriges kommuner och landsting. URL: <https://webbutik.skl.se/bilder/artiklar/pdf/7585-455-7.pdf?issuosl=ignore> (hämtad 2019-01-15).

Bilagor

Bilaga 1 - Forecastvärden enligt $SARIMA(1,0,1)(0,1,1)_{12}$ för Storstäder

	POINT.FORECAST	LO.95	HI.95	LO.99	HI.99	VERKLIGA
FEB 2016	0.012169	0.004695	0.019643	0.0023462	0.021991	0.015592
MAR 2016	0.015115	0.007168	0.023061	0.0046716	0.025558	0.019148
APR 2016	0.013504	0.005240	0.021769	0.0026429	0.024366	0.0139501
MAY 2016	0.020310	0.011827	0.028793	0.0091618	0.031459	0.0210625
JUN 2016	0.015384	0.006750	0.024019	0.0040371	0.026732	0.0191477
JUL 2016	0.014929	0.006189	0.023668	0.0034425	0.026415	0.0153181
AUG 2016	0.016733	0.007919	0.025547	0.0051499	0.028317	0.022775
SEP 2016	0.015831	0.006965	0.024696	0.0041788	0.027482	0.0276274
OCT 2016	0.016682	0.007780	0.025585	0.0049820	0.028382	0.0229772
NOV 2016	0.015783	0.006855	0.024712	0.0040493	0.027517	0.0142240
DEC 2016	0.014065	0.005118	0.023012	0.0023067	0.025823	0.0175065
JAN 2017	0.014120	0.005236	0.023155	0.0024208	0.025971	0.0182994
FEB 2017	0.011339	0.002123	0.020555	-0.000773	0.023451	0.0121099
MAR 2017	0.014417	0.005149	0.023684	0.0022373	0.026596	0.0137246
APR 2017	0.012917	0.003613	0.022220	0.0006900	0.025144	0.0153392
MAY 2017	0.019816	0.010487	0.029145	0.0075552	0.032076	0.0139937
JUN 2017	0.014968	0.005621	0.024315	0.0026839	0.027252	0.0153392
JUL 2017	0.014578	0.005218	0.023938	0.0022772	0.026879	0.0148010
AUG 2017	0.016438	0.007069	0.025807	0.0041255	0.028751	0.0139937
SEP 2017	0.015582	0.006207	0.024958	0.0032611	0.027904	0.0145319
OCT 2017	0.016473	0.007093	0.025853	0.0041459	0.028800	0.0158774
NOV 2017	0.015607	0.006224	0.024990	0.0032760	0.027939	0.0158774
DEC 2017	0.013917	0.004531	0.023302	0.0015824	0.026251	0.0113026
JAN 2018	0.014071	0.004684	0.023458	0.0017348	0.026408	
FEB 2018	0.011234	0.001657	0.020811	-0.001352	0.023821	
MAR 2018	0.014328	0.004721	0.023935	0.0017028	0.026954	
APR 2018	0.012843	0.003215	0.022470	0.0001893	0.025496	
MAY 2018	0.019753	0.010110	0.029396	0.0070803	0.032426	
JUN 2018	0.014915	0.005262	0.024569	0.0022288	0.027602	
JUL 2018	0.014534	0.004873	0.024195	0.0018374	0.027230	
AUG 2018	0.016401	0.006735	0.026067	0.0036977	0.029104	
SEP 2018	0.015551	0.005881	0.025221	0.0028427	0.028259	
OCT 2018	0.016447	0.006774	0.026119	0.0037349	0.029158	
NOV 2018	0.015585	0.005911	0.0252560	0.0028710	0.028299	
DEC 2018	0.013898	0.004222	0.0235737	0.0011821	0.026614	
JAN 2019	0.014056	0.004379	0.0237321	0.0013384	0.026773	
FEB 2019	0.011221	0.001366	0.0210759	-0.001731	0.024173	
MAR 2019	0.014317	0.004435	0.0241990	0.0013303	0.027304	
APR 2019	0.012833	0.002932	0.0227340	-0.000179	0.025845	
MAY 2019	0.019745	0.009831	0.0296595	0.0067157	0.032775	
JUN 2019	0.014909	0.004985	0.0248326	0.0018668	0.027951	
JUL 2019	0.014528	0.004598	0.0244588	0.0014773	0.027579	
AUG 2019	0.016396	0.006461	0.0263316	0.0033392	0.029453	
SEP 2019	0.015547	0.005608	0.0254856	0.0024854	0.028609	
OCT 2019	0.016443	0.006502	0.0263843	0.0033786	0.029508	
NOV 2019	0.015582	0.005640	0.0255251	0.0025155	0.028649	
DEC 2019	0.013896	0.003952	0.0238396	0.0008272	0.026964	
JAN 2020	0.014054	0.004109	0.0239983	0.0009839	0.027123	

Bilaga 2 – Forecastvärden enligt $SARIMA(1,0,1)(1,1,1)_{12}$ för Större städer

	POINT.FORECAST	LO.95	HI.95	LO.99	HI.99	VERKLIGA VÄRDEN
FEB 2016	0.010048	0.006170	0.0139263	0.004951	0.015145	0.014695
MAR 2016	0.010872	0.006918	0.0148258	0.005675	0.016068	0.014695
APR 2016	0.011711	0.007683	0.0157389	0.006417	0.017005	0.009184
MAY 2016	0.013769	0.009669	0.0178684	0.008381	0.019157	0.016531
JUN 2016	0.013080	0.008910	0.0172497	0.007600	0.018560	0.012333
JUL 2016	0.013970	0.009732	0.0182075	0.008400	0.019539	0.017581
AUG 2016	0.015008	0.010703	0.0193120	0.009350	0.020665	0.012858
SEP 2016	0.014166	0.009796	0.0185353	0.008423	0.019908	0.017581
OCT 2016	0.013085	0.008652	0.0175174	0.007259	0.018910	0.013120
NOV 2016	0.011294	0.006799	0.0157887	0.005386	0.017201	0.011021
DEC 2016	0.011591	0.007036	0.0161470	0.005604	0.017578	0.009184
JAN 2017	0.011519	0.006904	0.0161335	0.005455	0.017583	0.015278
FEB 2017	0.010099	0.005403	0.0147958	0.003927	0.016271	0.012429
MAR 2017	0.010996	0.006239	0.0157531	0.004744	0.017248	0.010099
APR 2017	0.011812	0.006995	0.0166291	0.005482	0.018143	0.013465
MAY 2017	0.014196	0.009321	0.0190712	0.007789	0.020603	0.011393
JUN 2017	0.013006	0.008073	0.0179382	0.006524	0.019488	0.012170
JUL 2017	0.014039	0.009051	0.0190278	0.007484	0.020595	0.014760
AUG 2017	0.015324	0.010281	0.0203677	0.008696	0.021952	0.017867
SEP 2017	0.014125	0.009028	0.0192219	0.007426	0.020824	0.017349
OCT 2017	0.013555	0.008405	0.0187050	0.006787	0.020323	0.012170
NOV 2017	0.011455	0.006253	0.0166566	0.004618	0.018291	0.012429
DEC 2017	0.011958	0.006705	0.0172103	0.005054	0.018861	0.010099
JAN 2018	0.011950	0.006648	0.0172529	0.004982	0.018919	
FEB 2018	0.010356	0.004935	0.0157773	0.003232	0.017481	
MAR 2018	0.011246	0.005766	0.0167262	0.004045	0.018448	
APR 2018	0.012063	0.006526	0.0176008	0.004786	0.019341	
MAY 2018	0.014423	0.008829	0.0200168	0.007071	0.021775	
JUN 2018	0.013267	0.007618	0.0189167	0.005843	0.020692	
JUL 2018	0.014290	0.008586	0.0199934	0.006793	0.021786	
AUG 2018	0.015556	0.009799	0.0213133	0.007989	0.023122	
SEP 2018	0.014381	0.008571	0.0201906	0.006745	0.022016	
OCT 2018	0.013774	0.007912	0.0196350	0.006070	0.021477	
NOV 2018	0.011694	0.005782	0.0176065	0.003924	0.019464	
DEC 2018	0.012181	0.006219	0.0181434	0.004346	0.020017	
JAN 2019	0.012169	0.006158	0.0181796	0.004269	0.020068	
FEB 2019	0.010586	0.004466	0.0167055	0.002543	0.018628	
MAR 2019	0.011475	0.005299	0.0176519	0.003358	0.019593	
APR 2019	0.012291	0.006059	0.0185236	0.004101	0.020482	
MAY 2019	0.014651	0.008364	0.0209386	0.006389	0.022914	
JUN 2019	0.013492	0.007151	0.0198335	0.005159	0.021826	
JUL 2019	0.014515	0.008121	0.0209085	0.006111	0.022918	
AUG 2019	0.015781	0.009335	0.0222273	0.007310	0.024253	
SEP 2019	0.014603	0.008106	0.0211006	0.006064	0.023142	
OCT 2019	0.013998	0.007450	0.0205454	0.005393	0.022603	
NOV 2019	0.011916	0.005319	0.0185133	0.003246	0.020586	
DEC 2019	0.012403	0.005757	0.0190493	0.003669	0.021138	
JAN 2020	0.012390	0.005696	0.0190840	0.003593	0.021187	

Bilaga 3 – Forecastvärden enligt $SARIMA(0,0,0)(0,1,1)_{12}$ för Mindre städer och landsbygd

	POINT.FORECAST	LO.95	HI.95	LO.99	HI.99	VERKLIGA VÄRDEN
FEB 2016	0.003818	0.000471	0.007166	-0.000581	0.008218	0.005933
MAR 2016	0.003546	0.000199	0.006894	-0.000853	0.007946	0.003955
APR 2016	0.005601	0.002253	0.008948	0.001201	0.010000	0.005537
MAY 2016	0.005196	0.001849	0.008544	0.000797	0.009596	0.004746
JUN 2016	0.005044	0.001696	0.008391	0.000645	0.009443	0.005537
JUL 2016	0.005975	0.002628	0.009323	0.001576	0.010374	0.008306
AUG 2016	0.005610	0.002263	0.008957	0.001211	0.010009	0.009888
SEP 2016	0.005320	0.001972	0.008667	0.000920	0.009719	0.006328
OCT 2016	0.004951	0.001603	0.008298	0.000551	0.009350	0.006724
NOV 2016	0.005136	0.001789	0.008483	0.0007368	0.009535	0.004351
DEC 2016	0.005402	0.002054	0.008749	0.0010024	0.009801	0.003955
JAN 2017	0.004582	0.001235	0.007930	0.0001831	0.008982	0.002360
FEB 2017	0.003818	0.000252	0.007384	-0.000868	0.008505	0.003933
MAR 2017	0.003546	-0.000020	0.007112	-0.001140	0.008233	0.005507
APR 2017	0.005601	0.002035	0.009167	0.000914	0.010287	0.007867
MAY 2017	0.005196	0.001630	0.008763	0.000510	0.009883	0.012980
JUN 2017	0.005044	0.001478	0.008610	0.0003571	0.009730	0.007473
JUL 2017	0.005975	0.002409	0.009541	0.001289	0.010662	0.006293
AUG 2017	0.005610	0.002044	0.009176	0.000923	0.010297	0.007080
SEP 2017	0.005320	0.001754	0.008886	0.000633	0.010006	0.005507
OCT 2017	0.004951	0.001384	0.008517	0.000264	0.009637	0.005507
NOV 2017	0.005136	0.001570	0.008702	0.000449	0.009823	0.005113
DEC 2017	0.005402	0.001836	0.008968	0.000715	0.010088	0.005113
JAN 2018	0.004582	0.001016	0.008148	-0.000104	0.009269	
FEB 2018	0.003818	0.000046	0.007590	-0.001139	0.008776	
MAR 2018	0.003546	-0.000226	0.007318	-0.001411	0.008504	
APR 2018	0.005601	0.001829	0.009373	0.000643	0.010558	
MAY 2018	0.005196	0.001424	0.008969	0.000239	0.010154	
JUN 2018	0.005044	0.001272	0.008816	0.000086	0.010001	
JUL 2018	0.005975	0.002203	0.009747	0.001018	0.010933	
AUG 2018	0.005610	0.001838	0.009382	0.000653	0.010567	
SEP 2018	0.005320	0.001548	0.009092	0.000362	0.010277	
OCT 2018	0.004951	0.001178	0.008723	-0.000007	0.009908	
NOV 2018	0.005136	0.001364	0.008908	0.000179	0.010094	
DEC 2018	0.005402	0.001630	0.009174	0.000444	0.010359	
JAN 2019	0.004582	0.000810	0.008355	-0.000375	0.009540	
FEB 2019	0.003818	-0.000149	0.007786	-0.001396	0.009033	
MAR 2019	0.003546	-0.000421	0.007514	-0.001668	0.008760	
APR 2019	0.005601	0.001633	0.009568	0.000386	0.010815	
MAY 2019	0.005196	0.001229	0.009164	-0.000002	0.010411	
JUN 2019	0.005044	0.001076	0.009011	-0.000170	0.010258	
JUL 2019	0.005975	0.002008	0.009943	0.0007609	0.011189	
AUG 2019	0.005610	0.001643	0.009578	0.0003959	0.010824	
SEP 2019	0.005320	0.001352	0.009287	0.0001055	0.010534	
OCT 2019	0.004951	0.000983	0.008918	-0.000264	0.010165	
NOV 2019	0.005136	0.001169	0.009104	-0.000078	0.010350	
DEC 2019	0.005402	0.001434	0.009369	0.000187	0.010616	
JAN 2020	0.004582	0.000615	0.008550	-0.000632	0.009797	