Prognoser för hotellmarknaden i Stockholm

Forecasts Concerning the Hotel Market in Stockholm

Linn Mattsson
Martin Wass

Handledare: Ann-Charlotte Hallberg
Examinator: Bertil Wegmann
Abstract

Background: This thesis targets hotels in Stockholm with aggregated data for the city. In the hotel market there’s three key indicators of particular interest and can be said describe how the market goes. Because of how much influence these key indicator have on the hotels it’s in great interest for the hotels to compare themselves with the market values. If these key indicators where forecasted it would perhaps be of great interest for the hotels to buy these forecasts to be able to control the room pricing in advance.

Purpose: Develop forecasting models due to future event and bookings with occupancy and room revenue as response variables. The key indicators revenue per available rooms and average price is then calculated through these forecasts for the year 2016.

Method: Since data consist of response variables (called output series) where the future values this series depends on past values of this series and a multiple set of related time series and external events (called input series) a dynamic regression called “regression with ARMA errors” where used. The method implies that you suit a multiple regression where the error terms are modelled with an appropriate ARMA model.

Results: The model for occupancy consist of four dependent variables and the model for the room revenue contain seven dependent variables. The error terms for these models include an autoregressive model with both seasonal and non-seasonal orders.

The forecast for the key indicators seems to follow the same pattern as previous years, where the event type Event more often than not gives a high estimate in relation to the current month. The event type Holiday seems to have a negative impact and bookings has a small positive effect for both models.

Conclusions: The models seems to fit data well but the input series needs more processing where the variable event seems to need some subgrouping. To forecast the room revenue is seems like a variable explaining price changes need to be constructed.
**Sammanfattning**

**Bakgrund:** Denna uppsats riktar in sig på hotell i Stockholm och all data som anges gäller för staden som helhet. Inom Hotellbranschen finns det tre vedertagna nyckeltal som kan sägas beskriva hur det går ekonomiskt för ett hotell. Då hotellen till stor del styrs efter dessa tre nyckeltal så är det av stort intresse för varje enskilt hotell att jämföra sina egna värden med marknadens värden på dessa nyckeltal. Om prognoser utförs på dessa nyckeltal borde det vara av stort intresse för varje hotell att ta del av dessa prognoser för att på så vis kunna reglera prissättningen uteftor hur marknaden kommer att se ut den närmaste tiden.

**Syfte:** Ta fram modeller som utifrån framtida evenemang och framtida bokningsläge prognostiserar hotellmarknadens Beläggning och Rumsintäkter. Utifrån dessa prognoser beräknas nyckeltalen Beläggning, Snittpris och intäkt per disponibelt rum på dagsnivå ett år fram i tiden, det vill säga för år 2016.

**Metod:** Då datamaterialet består av tidsserier med tillhörande förklarande variabler används en typ av dynamisk regressionsmodell. Dessa modeller är utformade för att hantera tidsseriedata med tillhörande förklarande variabler. Modellen som används kallas för regression med ARMA-fel och syftar till att en multipel regression anpassas och en lämplig ARMA-modell tas fram för att förklara feltermerna. På så vis modelleras även autokorrelationen som annars finns kvar i feltermerna.


Prognosen för nyckeltalen ser till största del ut att följa föregående års mönster, och evenemangstypen Event ger oftast en hög skattning i förhållande till månaden. Evenemangstypen Högtid tycks ge en negativ effekt och Bokningsläget har en positiv effekt för båda modellerna.

**Slutsats:** Modellerna anses välanpassade men det krävs mer bearbetning på de förklarande variablerna där till exempel event bör grupperas in beroende på vad för slags event det är. För att prognostisera rumsintäkter bör en variabel som förklara hotellens prisjusteringar modelleras.
Innehållsförteckning

1 Bakgrund................................................................................................................. 1
   1.1 Tidigare studier ............................................................................................... 1
   1.2 Syfte och frågeställningar ............................................................................ 2
   1.3 Etiska och samhälleliga aspekter .................................................................. 2
   1.4 Uppdragsgivare ............................................................................................. 2

2 Data ......................................................................................................................... 3
   2.1 Nyckeltal med tillhörande variabler ................................................................. 3
      2.1.1 Förra årets Beläggning ............................................................................ 6
   2.2 Evenemang ..................................................................................................... 6
      2.2.1 Uppdelning av evenemang ..................................................................... 7
   2.3 Bokningsläget .................................................................................................. 8
   2.4 Indikatorvariabler ........................................................................................... 9
   2.5 Rensning av förklarande variabler .................................................................. 11
      2.5.1 Förra Året ............................................................................................... 11
      2.5.2 Bokning ................................................................................................. 12
   2.6 Slutliga variabler ........................................................................................... 13

3 Metod ....................................................................................................................... 14
   3.1 Multipel regression ......................................................................................... 14
   3.2 Autokorrelation .............................................................................................. 14
   3.3 SAC och SPAC ............................................................................................... 14
   3.4 Conditional Least Squares ............................................................................. 15
   3.5 Dynamisk regression och ARMA ................................................................... 16
   3.6 Identifiering av lämplig ARMA-modell ............................................................ 17
   3.7 Variance Inflation Factor .............................................................................. 18
   3.8 Ljung-Box-test .............................................................................................. 18
   3.9 Justerad förklaringsgrad ............................................................................... 18

4 Resultat .................................................................................................................... 20
   4.1 Modell beläggning .......................................................................................... 20
      4.1.1 Modell 1 .................................................................................................. 20
      4.1.2 Modell 2 .................................................................................................. 23
   4.2 Modell rumsintäkter ...................................................................................... 26
      4.2.1 Modell 3 .................................................................................................. 27
      4.2.2 Modell 4 .................................................................................................. 30
   4.3 Prognoser ........................................................................................................ 33
4.3.1 Beläggning.............................................................................................................. 33
4.3.2 Rumsintäkter........................................................................................................... 34
4.3.3 Snittpris..................................................................................................................... 36
4.3.4 RevPAR..................................................................................................................... 37
5 Diskussion ..................................................................................................................... 40
  5.1 Datamaterialets lämplighet........................................................................................ 40
  5.2 Modellernas lämplighet.............................................................................................. 41
    5.2.1 Dynamisk regression............................................................................................ 41
    5.2.2 Beläggning............................................................................................................ 41
    5.2.3 Rumsintäkter......................................................................................................... 41
  5.3 Resultat ..................................................................................................................... 42
    5.3.1 Modell för beläggning........................................................................................... 42
    5.3.2 Modell för rumsintäkter ...................................................................................... 42
    5.3.3 Granskning av prognoser.................................................................................... 43
6 Slutsatser ....................................................................................................................... 44
7 Referenslista .................................................................................................................. 45
Bilaga 1 ............................................................................................................................. i
Bilaga 2 ............................................................................................................................. ii
Bilaga 3 ............................................................................................................................. iii
Bilaga 4 ............................................................................................................................. iv
Bilaga 5 ............................................................................................................................. v
1 Bakgrund


Det finns tre variabler, även kallade nyckeltal, som är extra intressanta vid analys av hotellmarknaden; beläggningen, snittpriset och intäkt per disponibelt rum. Ett hotell strävar ofta mot att hitta en balans för att ha höga värden på alla dessa nyckeltal samtidigt. Dessa tre nyckeltal medför att det också är naturligt att ett hotell vill kunna jämföra sig med andra hotell för att se hur de ligger till gentemot den allmänna hotellmarknaden. Då hotellen inte själva kan göra jämförelser sinsemellan utan att börja närma sig en kartellrörelse (Konkurrenslagen, SFS 2008:579) samlar vår uppdragsgivare Benchmarking Alliance in och sammanställer dessa data.


1.1 Tidigare studier


Det verkar vara ovanligt att försöka göra prognoser givet redan planerade händelser och på grund av detta finns det inget tydligt tillvägagångssätt för denna rapports prognostisering.

1.2 Syfte och frågeställningar

Uppsatsens syfte är att komma fram till en modell som klarar av att prognostisera nyckeltalen beläggnings, snittpris och intäkt per disponibelt rum så precis som möjligt med hänsyn tagen till framtida evenemang och bokningsläge. Våra prognoser gäller endast för år 2016 men förhoppningen är att modellen ska kunna ligga till grund för en förbättrad prognos även senare år.

Hotellen som ingår i studien är kunder till vår uppdragsgivare och omfattar cirka 100 större hotell i Stockholm. Frågeställningarna som ska undersökas är:

- Hur ser de slutliga modellerna för beläggningen och rumsintäkterna ut?
- Hur ser beläggningen ut på dagsnivå för hotellbranschen i Stockholm år 2016?
- Hur ser snittpriset per sält rum ut på dagsnivå för hotellbranschen i Stockholm år 2016?
- Vad är intäkten per disponibelt rum på dagsnivå för hotellbranschen i Stockholm år 2016?

1.3 Etiska och samhälleliga aspekter


1.4 Uppdragsgivare

2 Data

Uppsatsen utgår ifrån tre separata datamaterial som kommer från uppdragsgivaren där de antingen själva har samlad in data eller erhållit det inrapporterat från enskilda hotell. Samtliga datamaterial innehåller dagsdata. I tabell 1 visas samtliga variabler från rådata samt i vilket stycke dessa variabler beskrivs mer utförligt. I avsnitt 2.4 undersöks om utomstående indikatorvariabler ska inkluderas i prognoserna, och i avsnitt 2.5 rensas två förklarande variabler på icke-relevant varians. Kapitlet avslutas med en sammanfattande tabell för samtliga variabler som används i de slutliga modellerna.

Tabell 1: Sammanfattning variabler rådata.

<table>
<thead>
<tr>
<th>Variabelnamn</th>
<th>Förklaring</th>
<th>Stycke</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beläggning</td>
<td>Nyckeltal. Hur stor andel av hotellen som är uthyrt.</td>
<td>2.1</td>
</tr>
<tr>
<td>Snittpris</td>
<td>Nyckeltal. Det genomsnittliga priset per rum.</td>
<td>2.1</td>
</tr>
<tr>
<td>RevPAR</td>
<td>Nyckeltal. Intäkten per disponibelt rum.</td>
<td>2.1</td>
</tr>
<tr>
<td>SåldaRum</td>
<td>Antalet sålda rum.</td>
<td>2.1</td>
</tr>
<tr>
<td>Disponibla rum</td>
<td>Antalet tillgängliga rum för hotellmarknaden i Stockholm.</td>
<td>2.1</td>
</tr>
<tr>
<td>Rumsintäkter</td>
<td>Intäkterna för de sålda rummen för Stockholm som helhet.</td>
<td>2.1</td>
</tr>
<tr>
<td>Antal Besökare</td>
<td>Uppskattat antal besökare till evenemanget.</td>
<td>2.2</td>
</tr>
<tr>
<td>Andel från Sthlm</td>
<td>Uppskattad andel besökare som är från Stockholm. Kategorisk variabel, antar sju olika värden.</td>
<td>2.2</td>
</tr>
<tr>
<td>Marknad</td>
<td>Vilken marknad evenemanget påverkar. Antingen lokalt för Stockholm eller nationellt.</td>
<td>2.2</td>
</tr>
<tr>
<td>Regelbundenhet</td>
<td>Hur ofta evenemanget sker, både engångshändelser samt återkommande händelser. Kategorisk variabel.</td>
<td>2.2</td>
</tr>
<tr>
<td>Kategori</td>
<td>Visar om evenemanget riktar sig mot företagande eller fritid/hobbyverksamhet.</td>
<td>2.2</td>
</tr>
<tr>
<td>Bokning</td>
<td>Antalet bokade rum januari till november nästkommande år.</td>
<td>2.3</td>
</tr>
</tbody>
</table>

2.1 Nyckeltal med tillhörande variabler

Det mest omfattande datamaterialet är mellan åren 2008-2015 och innehåller aggregerad data för de cirka 100 hotellen i Stockholm. Det innehåller antalet sålda rum (SåldaRum), Rumsintäkter och Disponibla rum samt nyckeltalen Beläggning, Snittpris och intäkt per disponibelt rum (RevPAR, Revenue Per Available Room). Först visas hur de tre nyckeltalen beräknas i formel 1-3.

\[
Beläggning = \frac{SåldaRum}{Disponibla rum}
\]

\[
Snittpris = \frac{Rumsintäkter}{SåldaRum}
\]

\[
RevPAR = \frac{Rumsintäkter}{Disponibla rum}
\]


![Beläggning åren 2008-2015](image)

**Figur 1: Beläggning åren 2008-2015**

Figur 2: SåldaRum och Disponibla rum åren 2008-2015

Figur 2 visar att SåldaRum har en tydlig årsmönster där sommaren har en högre nivå och att variansen är lägre under dessa månader. Den positiva trenden de båda serierna har beror till största del på att fler hotell börjat använda vår uppdragsgivares tjänst och inte på att antalet hotell ökat nämnvärt från och med år 2008.

Disponibla rum lägger sig som ett tak över sålda rum och ökar då fler hotell ansluter sig till tjänsten, men även när nya hotell öppnas. De små nedgångarna som syns för Disponibla rum beror på att hotellen tillfälligt kan stänga rum för exempelvis renovering, vilket gör att dessa rum räknas som otillgängliga under denna period.
I figur 3 utläsas att **Rumsintäkter** har en månadsvariation där sommarmånadernas nivå är högre än övriga månader och att det följer ett mönster som liknar det för **SäldaRum**. Det finns även återkommande extrempunkter, en till två grupper av efterföljande dagar om året.

Det finns utöver detta även information om hotell som förväntas öppna under år 2016, som kommer användas för att uppskatta **Disponibla rum**.

2.1.1 Förra årets Beläggning

Utifrån månadsvariation för **Beläggning** och eftersom även veckodagen påverkar (se bilaga 3) valde vi att även använda en variabel som jämför **Beläggning** detta år med förra året samma vecka, samma veckodag. Till exempel så jämförs tisdag vecka 4, 2015 med tisdag vecka 4, 2014, där observationen år 2014 kallas för **Beläggning förra året** (förkortas **FörraÅret**). Denna variabel, som kan sägas är tidsförskjutning 364 av **Beläggning**, har en korrelation på 0,832 med **Beläggning** och är därför lämplig att använda som förklarande variabel.

2.2 Evenemang


Endast en av dessa fem variabler, **Marknad**, finns inlagt kontinuerligt för samtliga år. De fyra övriga variablerna saknas för nästan alla evenemang fram till år 2014. Vi skapade även ytterligare en

### 2.2.1 Uppdelning av evenemang


*Tabell 2: Korrelation mellan evenemangsvariablerna mot SåldaRum samt Beläggning.*

<table>
<thead>
<tr>
<th></th>
<th>Antal Besökare</th>
<th>Andel från Sthlm</th>
<th>Antal utifrån Sthlm</th>
<th>Marknad</th>
<th>Regelbundenhet</th>
<th>Kategori</th>
</tr>
</thead>
<tbody>
<tr>
<td>SåldaRum</td>
<td>0,011</td>
<td>-0,173</td>
<td>0,015</td>
<td>0,399</td>
<td>0,031</td>
<td>0,036</td>
</tr>
<tr>
<td>Beläggning</td>
<td>0,012</td>
<td>-0,18</td>
<td>0,016</td>
<td>0,388</td>
<td>0,030</td>
<td>0,039</td>
</tr>
</tbody>
</table>

Det utläses att det endast är två variabler som har korrelationer med absolutvärden större än 0,05; *Andel från Sthlm* samt *Marknad*. Variabeln *Marknad* undersöktes först då den hade högst korrelation med båda variablerna. *Marknad* kan anta tre olika värden, där ett av värdena motsvarar nationella evenemang (högtider) och de två andra betecknar evenemang i Stockholmsområdet. Totalt är det 738 evenemang på åtta år.

I figur 4 visas *SåldaRum* och *Beläggning* plottade mot varandra med evenemangen grupperade i högtider, övriga event och när båda evenemangstyperna sker samtidigt.

![SåldaRum och beläggning uppdelat efter eventtyp](image)

*Figur 4: SåldaRum mot Beläggning grupperat på Event, Högtider och EventHögtider.*

Figur 4 visar att grupperna överlappar varandra, men de högsta värdena för både *SåldaRum* och *Beläggning* är övriga event och de lägsta värdena är högtider. Gruppen för båda evenemangstyperna hamnar centralt på linjen. Utifrån den relativt höga korrelationen samt mönstret i figur 4 anses denna uppdelning av evenemangen lämplig. Det kan även motiveras logiskt, som att exempelvis jul och påsk är helgdagar som de flesta firar hemma till skillnad från övriga event och att evenemangen
därför borde delas upp på detta. Då evenemangen fortsättningsvis alltid är uppdelade i dessa 
undergrupper benämns hädanefter Övriga Event endast som Event, högtiderna kallas för Högtider 
och gruppen båda evenemangstyperna kallas för EventHögtider (förkortas EH).

Även Andel från Sthlm undersöcktes, men ett stort problem är det stora antalet saknade värden. Av 
totalt 1033 evenemangsdagar under de åtta åren har endast 430 evenemangsdagar värden på 
andelen från Stockholm. Anledningen till att det är fler evenemangsdagar än de 738 evenemangen är 
eftersom de individuella evenemangen kan vara längre än en dag. Andel från Sthlm är en kategorisk 
variabel som kan anta sju olika värden och fördelningen på denna variabel visas i tabell 3.

Tabell 3: Fördelning av uppskattad andel besökare från Stockholm.

<table>
<thead>
<tr>
<th>Andel från Sthlm</th>
<th>0,05</th>
<th>0,25</th>
<th>0,50</th>
<th>0,75</th>
<th>0,80</th>
<th>0,90</th>
<th>1,00</th>
</tr>
</thead>
<tbody>
<tr>
<td>Antal observationer</td>
<td>72</td>
<td>6</td>
<td>25</td>
<td>66</td>
<td>2</td>
<td>218</td>
<td>41</td>
</tr>
</tbody>
</table>

Nästan hälften av observationerna har 90 % av besökarna från Stockholm och majoriteten av evenemangen har mer än 50 % av från Stockholm. När SäldeRum och Beläggning plottas grupperade på 
dessa andelar (se bilaga 4) är det enda tydliga mönstret att andelen 1,00, det vill säga när 100 % av 
besökarna uppskattas vara från Stockholm, ger höga värden på båda variablerna. När alla besökare 
på evenemanget är från Stockholm borde det snarare ge låga värden, vilket går emot det logiska. Då 
grafen inte visar förslag på någon alternativ uppdelning för andelen än denna beslutas att variabeln 
andelen besökare från Stockholm inte används för att gruppera evenemangen.

2.3 Bokningsläget

Den tredje datamängden består av bokningsläget för de nästkommande 365 dagarna. Hotellen har 
för åren 2012-2015 rapporterat det framtida bokningsläget i början av varje månad. Detta innebär 
att för bokningar inrapporterade i exempelvis oktober sträcker sig bokningarna från och med första 
oktober till och med sista september nästkommande år. Vi kommer att använda oss av bokningsläget 
inrapporterat i december då det ger bokningsläget för nästkommande år. Detta innebär att det 
första året det finns observationer för är år 2013. Bokningsläget visas visuellt i figur 5.

Det utläses att det finns fler bokade rum i början av året samt under sommaren. Antalet bokade rum är som lägst för årets sista månader, och som visas saknas värden för december varje år. Anledningen till att varje observerat år har en nedåtgående trend är för att bokningsläget från december används. Detta leder till att januari ligger mycket närmare i tiden än exempelvis augusti, vilket innebär att fler personer har hunnit bokat hotellrum för dessa dagar. I övrigt syns återkommande extrempunkter där många hotellrum redan är uppbokade trots att de ligger långt fram i tiden.

Kort sagt består bokningsläget av observationer från januari till och med november för åren 2013-2016, med saknade värden i december för samtliga år. Variabeln för bokningsläget kallas fortsättningsvis för Bokning.

2.4 Indikatorvariabler

Förutom evenemangsvariablerna används tre ytterligare indikatorvariabler i prognoserna. Dessa kallas för Månad, Helg och Juli.

De visuella analyserna av figur 1-5 visar på att flertalet variabler varierar över året. För att kunna ta hänsyn till detta skapas indikatorvariabeln Månad, som helt enkelt anger vilken månad observationen ligger i. Månaderna kodas 1 för januari fram till 12 för december.

Variabeln Helg utmärkte sig när sambandet mellan SäldaRum och Rumsintäkter undersöktes, se figur 6.
Figur 6: Sambandet mellan rumsintäkter och SåldaRum grupperat på Helg och vardag.

När observationerna delas in helgdagar (fredag-söndag) och vardagar (måndag-torsdag) som i figur 6 syns ett tydligt samband mellan de två grupperna. Helgdagar har generellt sett lägre rumsintäkter jämfört med vardagar trots att lika många rum har sålts. Det finns även en antydning om en tredje grupp som ligger högre än både vardagar och helgdagar. Utifrån grupperingen på Helg är det dock mer troligt att den gruppen består av extremvärden. Kodningen för Helg är 1 för helgdagar och 0 för vardagar, och variabeln Helg används för att förklara prognosen för Rumsintäkter.

I figur 6 syns även att helgdagar och vardagar har olika lutning. För att ta hänsyn till detta skapas en interaktionsterm mellan SåldaRum och Helg, kallad SåldaRum*Helg.

Till sist undersöktes Juli i figur 7 och även denna variabel har en ganska tydlig skillnad mellan SåldaRum och Rumsintäkter.
Figur 7: (a) SäldaRum grupperat på månad, (b) Rumsintäkter grupperat på månad.

SäldaRum och Rumsintäkter har i stort sett samma årsmönster, men månaden juli skiljer sig mellan de båda. I juli säljs ungefär lika många rum som i de närliggande månaderna, men Rumsintäkter är lägre. De övriga månaderna har i stort sett samma mönster för båda variablerna. Då nedgången i juli inte förklaras av SäldaRum används Juli för att indikera denna månad i prognosen för Rumsintäkter.

2.5 Rensning av förklarande variabler


2.5.1 FörraÅret

Variationen ser ut att ha minskat för R.FörraÅret (figur 8b) jämfört med den icke-rensade serien i figur 8a, men det är svårt att säga om det är en betydande skillnad eftersom skalan på y-axeln är olika på de båda graferna. Eftersom vi rensar på Event, som ofta ger toppar, och Högtider, som ofta ger dalar, är det logiskt att variansen minskar när vi rensar på de mest extrema värdena. Årsmönstret är i stort sett fortfarande samma för de båda serierna, vilket det också ska vara eftersom vi inte rensar för detta. Rent grafiskt ser det ut som att rensningen har gett gott resultat, då variansen minskat men det huvudsakliga mönstret finns kvar i data.

2.5.2 Bokning

Figur 9 visar att **Månad** har lyckats att justera upp **Bokning** för de senare månaderna och den nedåtgående trenden har försvunnit för den rensade serien **R.Bokning**. Variansen ser ut att vara oförändrad mellan de båda serierna och även årsmönstret är samma som innan. De högsta topparna utmärker sig dock fortfarande, vilket skulle kunna betyda att de tre evenemangsvariablerna inte påverkar **Bokning** lika mycket som det påverkar **FörraÅret**. Även denna rensning har till synes gett ett önskvärt resultat och variabeln **R.Bokning** används vidare i analysen.

### 2.6 Slutliga variabler

I tabell 4 sammanfattas variablerna som används i de kommande modellerna. Statusen **Prognostiseras** innebär att en modell anpassas för att få fram värdena för år 2016, medan **Beräknas** betyder att variabeln beräknas fram antingen genom tidigare kända värden eller genom prognoserna. Om statusen är **Känd** finns faktiska värden på variabeln år 2016, antingen direkt i rådata eller skapade utifrån rådata.

**Tabell 4: Sammanfattning variabler för modellerna.**

<table>
<thead>
<tr>
<th>Variabelnamn</th>
<th>Förklaring</th>
<th>Status år 2016</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beläggning</td>
<td>Nyckeltal. Hur stor andel av hotellen som är uthyr.</td>
<td>Prognostiseras</td>
</tr>
<tr>
<td>Snittpris</td>
<td>Nyckeltal. Det genomsnittliga priset per rum.</td>
<td>Beräknas</td>
</tr>
<tr>
<td>RevPAR</td>
<td>Nyckeltal. Intäkten per disponibelt rum.</td>
<td>Beräknas</td>
</tr>
<tr>
<td>SåldaRum</td>
<td>Antalet sålda rum.</td>
<td>Beräknas</td>
</tr>
<tr>
<td>Disponibla rum</td>
<td>Antalet tillgängliga rum för hotellmerknaden i Stockholm.</td>
<td>Känd</td>
</tr>
<tr>
<td>Rumsintäkter</td>
<td>Intäkterna för de sålda rummen för Stockholm som helhet.</td>
<td>Prognostiseras</td>
</tr>
<tr>
<td>Event</td>
<td>Evenemang som inte är högtider, t.ex. konsert eller mässor.</td>
<td>Känd</td>
</tr>
<tr>
<td>Högtider</td>
<td>Evenemang som är högtider, t.ex. påsk och jul.</td>
<td>Känd</td>
</tr>
<tr>
<td>EH</td>
<td>Dagar där Event och Högtider inträffar samtidigt.</td>
<td>Känd</td>
</tr>
<tr>
<td>Månad</td>
<td>Indikerar vilken månad på året det är.</td>
<td>Känd</td>
</tr>
<tr>
<td>Helg</td>
<td>Indikatorvariabel för dagarna fredag - söndag.</td>
<td>Känd</td>
</tr>
<tr>
<td>SåldaRum*Helg</td>
<td>Interaktionterm mellan SåldaRum och Helg.</td>
<td>Känd</td>
</tr>
<tr>
<td>Juli</td>
<td>Indikatorvariabel för månaden juli.</td>
<td>Känd</td>
</tr>
<tr>
<td>FörraÅret</td>
<td>Hur stor andel av hotellen som var uthyr 364 dagar tidigare.</td>
<td>Känd</td>
</tr>
<tr>
<td>Bokning</td>
<td>Antalet bokade rum januari till november nästkommende år.</td>
<td>Känd</td>
</tr>
<tr>
<td>R.FörraÅret</td>
<td>FörraÅret rensad för oönskad variation. Skapas.</td>
<td>Känd</td>
</tr>
<tr>
<td>R.Bokning</td>
<td>Bokning rensad för oönskad variation. Skapas.</td>
<td>Känd</td>
</tr>
</tbody>
</table>
3 Metod

De olika metoderna som används i uppsatsen tas upp i detta kapitel. För prognoserna för Beläggning och Rumsintäkter används en typ av dynamisk regressionsmodell.


För samtliga analyser har programmet SAS använts, och grafer har skapats i både SAS och R.

3.1 Multipel regression

En regressionsmodell som innehåller mer än en oberoende variabel kallas för en multipel regression. I formel 4 visas ekvationen för en generell multipel regression där $x_1, \ldots, x_{p-1}$ är de förklarande variablerna och $n$ är en felterm. (Bowerman, O’Connell & Koehler, 2005).

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_{p-1} x_{p-1} + n$$ (4)

där

$$n \sim iid \ N(0, \sigma^2)$$ (5)

Feltermerna $n$ antas vara oberoende och normalfördelade med medelvärde noll och varians $\sigma^2$.

3.2 Autokorrelation

Feltermerna, även kallade residualer, anses ha autokorrelation om tidigare värden på feltermen påverkar det nuvarande värdet på feltermen. Vid positiv autokorrelation tenderar en positiv felterm att följas upp av en positiv felterm och en negativ felterm tenderar att följas upp av en negativ felterm. Om positiva felterm följs upp av negativa och vice versa så antas en negativ autokorrelation. (Bowerman, O’Connell & Koehler, 2005).

Autokorrelation kan liknas med korrelationen mellan två variabler men att man studerar hur variabeln vid tidpunkten $t$ korrelerar med sig själv vid olika tidsförskjutningar $t-k$, där $k$ är antalet tidpunkter tillbaka i tiden.

3.3 SAC och SPAC

Autokorrelationsfunktionen (SAC) och den partiella autokorrelationsfunktionen (SPAC) används för att avgöra vilken struktur som feltermerna till tidsserieobservationer kan tänkas följa.

Med en spik menar man en tidsförskjutning vars autokorrelation har större absolutbelopp än t-statistikan och kan utläsas då SAC och SPAC illustreras i en graf. SAC mäter det linjära förhållandet mellan tidsserieobservationer med $k$ tidsenheter mellan sig. Formel 6 visar, utifrån tidsserievärdena $y_b, \ldots, y_n$, hur det kvantitativa måttet beräknas. Formel 6 och 7 visar hur måttenheten för SAC beräknas.

$$r_k = \frac{\sum_{t=b}^{n-k}(y_t - \bar{y})(y_{t+k} - \bar{y})}{\sum_{t=b}^{n}(y_t - \bar{y})^2}$$ (6)

där

$$\bar{y} = \frac{\sum_{t=b}^{n}y_t}{n - b + 1}$$ (7)
Standardavvikelsen för SAC beräknas enligt formel 8.

\[ s_{rk} = \begin{cases} 
\frac{1}{(n-b+1)^{1/2}} & \text{om } k = 1 \\
\frac{(1 + 2 \sum_{j=1}^{k-1} r_j^2)^2}{(n-b+1)^{1/2}} & \text{om } k = 2,3, \ldots
\end{cases} \tag{8} \]

t-statistiskan för SAC beräknas som formel 9 visar.

\[ t_{rk} = \frac{r_k}{s_{rk}} \tag{9} \]

\( r_k \) antar bara värden mellan -1 till 1 där ett värde nära 1 indikerar att observationer med \( k \) tidsenheter mellan sig har ett samband med i form av en linjär form med en positiv lutning. Ett \( r_k \) nära -1 tyder på ett starkt negativt samband mellan observationer med \( k \) tidsenheter mellan sig. SAC är en lista eller graf över autokorrelationen för tidsförskjutningarna \( k = 1,2, \ldots \). (Bowerman, O’Connell & Koehler, 2005).

Måttenheten för SPAC tolkas på samma sätt som SAC som förhållande mellan observationer med \( k \) tidsenheter mellan sig då effekterna av de mellanliggande observationerna eliminerade. SPAC beräknas enligt formel 10 och 11.

\[ r_k = \begin{cases} 
\frac{r_k}{\sum_{j=1}^{k-1} r_{k-1,j} r_{k-j}} & \text{om } k = 1 \\
\frac{1 - \sum_{j=1}^{k-1} r_{k-1,j} r_{k-j}}{1 - \sum_{j=1}^{k-1} r_{k-1,j} r_{j}} & \text{om } k = 2,3, \ldots
\end{cases} \tag{10} \]

där

\[ r_{kj} = r_{k-1,j} - r_{kk} r_{k-1,k-j} \quad \text{för } j = 1,2, \ldots, k - 1 \tag{11} \]

Standardavvikelsen och t-statistiskan för SPAC beräknas enligt formel 12 respektive formel 13.

\[ s_{rkk} = \frac{1}{(n-b+1)^{2}} \tag{12} \]

\[ t_{rkk} = \frac{s_{rkk}}{s_{rkk}} \tag{13} \]

SPAC är en lista eller graf över den partiella autokorrelationen för tidsförskjutningarna \( k = 1,2, \ldots \).  

3.4 Conditional Least Squares

Conditional Least Squares (CLS) är en skattningsmetod för ARIMA-modeller som antar att feltermar för tidigare observationer är noll, och är därför användbara vid rena AR-modeller. Tidsserien \( n_t \) kan skrivas som en funktion av tidigare observationer enligt formel 14.

\[ n_t = \alpha_t + \sum_{i=1}^{\infty} \pi_{i} n_{t-i} \tag{14} \]

Vikten \( \pi \) beräknas som kvoten mellan polynomen för \( \Phi \) och \( \Theta \) som visas i formel 15.

\[ \frac{\Phi(B)}{\Theta(B)} = 1 - \sum_{i=1}^{\infty} \pi_{i} B^i \tag{15} \]

CLS-metoden skattar koefficienterna genom att minimera formel 16.
Tidigare icke-observerade värden på $n_t$ sätts till noll och $\hat{\theta}_t$ beräknas utifrån skattningar för $\Phi$ och $\theta$ vid varje iteration.

### 3.5 Dynamisk regression och ARMA

Dynamiska regressionsmodeller är ett samlingsnamn för olika ARMA-modeller som klarar av att ta hänsyn till en eller flera förklarande tidsserier. I denna uppsats används den modell som anpassar en multipel linjär regression och samtidigt modellerar feltermerna $n_t$ från denna med en ARMA. När vi i fortsättningen nämner dynamisk regression är det denna modell som menas. I formel 21 visas ekvationen för en multipel linjär regression för en tidsserie. $x_{1,t}, \ldots, x_{p-1,t}$ är de förklarande variablerna vid tidpunkt $t$.

$$y_t = \beta_0 + \beta_1 x_{1,t} + \beta_2 x_{2,t} + \cdots + \beta_{p-1} x_{p-1,t} + n_t$$

Anledningen till att även feltermerna modelleras är eftersom när en linjär regression anpassas för en tidsserie blir feltermerna oftast inte oberoende. På grund av detta bör Box-Jenkins metodik användas (Bowerman, O’Connell & Koehler, 2005), ofta i form av en ARMA eller en ARIMA, för att hitta en modell som beskriver feltermerna.

ARMA-modeller är ett av de mest vedertagna tillvägagångssättet vid tidsserieprognoser (Hyndman & Athanasopoulos, 2012). Denna typ av modeller riktar in sig på att beskriva datamaterialets autokorrelationer till skillnad från exempelvis exponentiell utjämning som fokuserar på att beskriva trend och säsongsmonstret i data. ARMA($p$, $q$) är en förkortning för AutoRegressive Moving Average och kan sägas bestå av två fristående modeller; autoregressiva modeller (AR($p$)-modeller) och glidande medelvärdesmodeller (MA($q$)-modeller). $p$ och $q$ är parametrar som visar hur många AR-termer respektive MA-termer som modellen ska ta hänsyn till på icke-säsongsnivå. Prognoserna i AR-modellen bygger på att variabeln som prognostiseras kan skrivas som en linjärkombination av en eller flera tidigareobservationer av variabeln (formel 18). MA-modellen fungerar på ett liknande sätt, men istället för att använda sig av värden på tidigare observationerna används de tidigare feltermerna, i detta sammanhang även kallat vitt brus, för att göra prognoser på kommande värden (formel 19). I båda formlerna betecknar $B^k$ bakåtskiftsoperatorn som subtraheras indexet på $n_t$ med $k$ (se formel 20) och $a_t$ är vitt brus, som ska vara både stationärt samt oberoende. För ett alternativt sätt att skriva detta krav på $a_t$, se formel 21.

$$n_t = \delta + (\phi_1 B + \phi_2 B^2 + \cdots + \phi_p B^p) n_t + a_t$$

$$n_t = \delta + (1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q) a_t$$

$$B^k n_t = n_{t-k}$$

$$a_t \sim iid N(0, \sigma^2)$$

För att sedan avgöra om AR- och/eller MA-modellen ska användas samt vilken ordning av respektive modell som är lämplig studeras de första tidsförskjutningarna i SAC och SPAC för datamaterialet.

Det är också möjligt att ta hänsyn till säsongen med en säsongsARMA(SARMA) genom att lägga till en säsongsdel som anpassar AR- och MA-modeller på säsongsänd L, se formel 22 och 23. Modellen kan då skrivas SARMA($p$, $q$)($P$, $Q$), där $p$ och $q$ som tidigare nämnt hör till icke-säsong. $P$ och $Q$ hör till säsongsdelen och betecknar antalet SAR-termer respektive SMA-termer för säsongsänden L.
Ordningen av dessa bestäms liksom för ARMA utan säsong genom att undersöka datamaterialets SAC och SPAC. Riktlinjerna för vilken ARMA-modell som är lämplig kan hittas exempelvis i Bowerman, O’Connell & Koehler (2005).

\[ n_t = \delta + (\phi_1 B + \phi_2 B^2 + \cdots + \phi_p B^p) n_t + a_t \]  
\( (22) \)

\[ n_t = \delta + (1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q) a_t \]  
\( (23) \)


\[ y_t = \beta_0 + \beta_1 x_{1,t} + \beta_2 x_{2,t} + \cdots + \beta_p x_{p-1,t} + n_t \]  
\( (24) \)

\[ (1 - \phi_1 B - \cdots - \phi_p B^p)(1 - \phi_1 B - \cdots - \phi_p B^p) n_t = \delta + a_t \]  
\( (25) \)

\[ a_t \sim iid N(0, \sigma^2) \]  
\( (26) \)

I denna uppsats utförs prognoserna i form av två dynamiska regressioner som båda prognostisar Beläggning. Den första modellen för Beläggning använder sig utav \( R.\text{FörraÅret}, R.\text{Bokning}, Högtider, Event \) och \( EH \) som förklarande variabler. Eftersom \( R.\text{Bokning} \) innehåller saknade värden för åren 2009-2012 samt för december samtliga år klarar modellen endast av att ta hänsyn till de tre sista åren. För att få in ytterligare information om modeller december tillämpas andra modeller. För att få in ytterligare information om data samt för att lyckas modella december tillämpas ytterligare en dynamisk regression utan \( R.\text{Bokning} \) som förklarande variabel. Denna regression innehåller alltså endast \( R.\text{FörraÅret}, Högtider, Event \) och \( EH \) som förklarande variabler. Utifrån dessa två dynamiska modeller utförs en prognos per modell och de prognostiserade värdena från modell 2 imputeras till de saknade värdena för prognosen med den första modellen.

Eftersom Beläggning används som responsvariabel kontrolleras prognosen automatiskt för Disponibla rum eftersom Beläggning är en funktion av Disponibla rum (se formel 1). Uppskattningen av Disponibla rum för år 2016 används vidare för att korrigerar eventuella överskridningar i prognosen, då det är orimligt att \( SåldaRum \) överstiger Disponibla rum.

Utifrån prognosen över Beläggning tas Rumsintäkter för år 2016 fram. Då Disponibla rum ökade med nästan 50 % fram till år 2012 betyder det att Rumsintäkter kommer att vara lägre dessa år trots att Beläggning har samma värde eftersom färre rum har sälts. Både Beläggning och SåldaRum har i stort sett samma korrelation med Rumsintäkter (0,868 respektive 0,862) vilket innebär att båda variablerna är lämpliga som förklarande variabler. Då SåldaRum anses vara den variabel som har en större logisk koppling till Rumsintäkter och eftersom vi kan omvandla Beläggning till SåldaRum genom formel 1 väljer vi därför att sätta SåldaRum som förklarade variabel till Rumsintäkter.

Liksom för Beläggning anpassas två olika modeller för Rumsintäkter, en med \( R.\text{Bokning} \) och en utan. Förutom SåldaRum och R.\text{Bokning} förklaras också Rumsintäkter av Event, Högtider, EH, Helg, Juli och interactionstermen SåldaRum*Helg. Interactionstermen läggs till för att ta hänsyn till att de två Helggrupperna har olika lutning när SåldaRum och Rumsintäkter grupperas på detta, tidigare visat i figur 6.

### 3.6 Identifiering av lämplig ARMA-modell

För att identifiera strukturen för \( p \) och \( q \) så studeras SAC och SPAC på icke-säsongstidsförskjutningar. Om SAC tycks avta samtidigt som SPAC uppvisar spikar då alla andra tidsförskjutningar tycks ha lägre
$r_k$ jämfört med t-statistik kan så antas en AR(k)-modell där $k$ är den tidsförskjutning där den sista spiken gick att utläsa. (Bowerman, O’Connell & Koehler, 2005).

Vidare studeras SAC och SPAC på säsongstidsförskjutningarna $L$, $2L$, ... för att identifiera en lämplig struktur för $P$ och $Q$. Om SAC uppvissar ett avtagande mönster för var $L$:e tidsförskjutning och SPAC uppvissar spikar var $L$:e tidsförskjutning så antas en AR(k)-struktur för säsongsdelen. (Bowerman, O’Connell & Koehler, 2005). Modellen kan vidare skrivas som $SARMA(k, 0)(k, 0)_L$.

### 3.7 Variance Inflation Factor

Ett sätt att upptäcka multikolinjäritetsproblem i en regression är undersöka variablernas Variance Inflation Factor, även kallat VIF (Kutner et al., 2005). Hur VIF beräknas visas i formel 27.

\[
(VIF)_k = \frac{1}{1 - R_k^2}
\]

Genom att VIF tar hänsyn till förklaringsgraden för variablen $x_k$ visar VIF om $x_k$ har ett linjärt samband till de övriga förklarande variablerna. Det lägsta värdet VIF kan anta är ett, vilket innebär att $x_k$ inte har något linjärt samband med det övriga. Ju högre VIF desto mer korrelerar $x_k$ med det övriga, och en tumregel är att VIF över 10 tyder på att det finns multikolinjäritetsproblem (Kutner et al., 2005).

### 3.8 Ljung-Box-test

När tidsserier anpassas finns risk för att autokorrelation finns kvar i residualerna från modellen. För att undersöka detta kan ett Ljung-Box-test utföras. Testet undersöker om modellen kan anses vara bristfällig på grund av att autokorrelation i residualerna kvarstår. Hypotesererna för testet visas och hur teststatistikan beräknas syns i formel 28. $n$ står för antalet observationer för responsvariabeln, $\hat{r}_k^2$ är den skattade autokorrelationen på tidsförskjutning $k$, och $m$ är antalet tidsförskjutningar som testet undersöker (NIST/SEMATECH, 2013).

**$H_0$:** Residualerna har ingen autokorrelation.

**$H_a$:** Residualerna har autokorrelation.

\[
Q = n(n + 2) \sum_{k=1}^{m} \frac{\hat{r}_k^2}{n - k}
\]

Om $Q > \chi^2_{1-a,m-p-q}$ förkastas $H_0$ på signifikansnivån $\alpha$. Tabellvärdet $\chi^2_{1-a,m-p-q}$ är $\chi^2$-fördelad med $m - p - q$ frihetsgrader, där där $p$ och $q$ kommer från antalet parametrar från den undersöka modellen ARMA($p$, $q$). När $H_0$ kan förkastas gäller det att residualerna har autokorrelation och modellen anses därför vara bristfällig. Till skillnad från de flesta andra test anses modellen alltså lämplig när $H_0$ inte kan förkastas.

### 3.9 Justerad förklaringsgrad

Den icke-justerade förklaringsgraden ($R^2$) visar hur stor del av variationen som förklaras av modellens oberoende variabler. Denna typ av förklaringsgrad ökar alltid när ytterligare oberoende variabler läggs till, eftersom även olämpliga variabler marginellt kan förklara variationen. Den justerade förklaringsgraden ($\tilde{R}^2$) tar hänsyn till hur bra responsvariabeln förklaras med hänsyn till hur många oberoende variabler modellen har. Det är alltså möjligt att $\tilde{R}^2$ minskar när modellen får fler oberoende variabler. Hur $R^2$ och $\tilde{R}^2$ beräknas syns i formel 29 och 30 (SAS Institute Inc., u.å.(b)).
\[ R^2 = 1 - \frac{SS_{error}}{SS_{total}} \]  

\[ \bar{R}^2 = 1 - \frac{(n - 1)(1 - R^2)}{n - p} \]  

Den icke-justerade förklaringsgraden används oftast för att utvärdera den individuella modellens lämpliga och den justerade förklaringsgraden används främst för att jämföra olika modeller. I våra analyser vill vi se vilken påverkan variabeln \textit{R.Bokning} har på modellen, och vi väljer därför att studera den justerade förklaringsgraden för samtliga modeller.
4 Resultat


4.1 Modell beläggning

Beläggning prognostiseras med hjälp av två separata dynamiska regressioner. Som tidigare nämnt består den ena modellen av $R.\text{FörraÅret}$, $R.\text{Bokning}$ samt de tre evenemangsvariablerna, och den andra modellen består av $R.\text{FörraÅret}$ tillsammans med evenemangsvariablerna. Regressions- och ARMA-parametrarna skattas sedan samtidigt för att optimera funktionen. I tabell 5 visas en sammanfattning över de variabler som används till dessa modeller.

Tabell 5: Förklarande variabler, Beläggning.

<table>
<thead>
<tr>
<th>Variabelnamn</th>
<th>Förklaring</th>
<th>Status år 2016</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beläggningen</td>
<td>Hur stor andel av hotellen som är uthyrt.</td>
<td>Prognostiseras</td>
</tr>
<tr>
<td>R.\text{FörraÅret}</td>
<td>FörraÅret justerat för oönskad variation. Skapad.</td>
<td>Känd</td>
</tr>
<tr>
<td>R.\text{Bokning}</td>
<td>Bokning justerat för oönskad variation. Skapad</td>
<td>Känd</td>
</tr>
<tr>
<td>Event</td>
<td>Evenemang som inte är högtider, t.ex. konserter eller mässor.</td>
<td>Känd</td>
</tr>
<tr>
<td>Högtider</td>
<td>Evenemang som är högtider, t.ex. påsk och jul.</td>
<td>Känd</td>
</tr>
<tr>
<td>EH</td>
<td>Dagar där Event och Högtider inträffar samtidigt.</td>
<td>Känd</td>
</tr>
</tbody>
</table>

Till att börja med anpassas den multipla linjära regressionen och SAC och SPAC för residualerna studeras. Utifrån detta identifieras en lämplig ARMA-modell till den dynamiska regressionen. Vidare skattas den multipla regression och den valda modellen för $n_t$ samtidigt I form av en dynamisk regression. $a_t$ för den dynamiska regressionen studeras både visuellt och genom ett Ljung-Box-test för att hitta den mest lämpade ARMA-modell. Om det fortfarande finns autokorrelation kvar anpassas en ny modell utifrån vad som visas i SAC och SPAC. Detta upprepas tills en lämplig modell är framtagen.

4.1.1 Modell 1

Första steget i den dynamiska regressionen till modellen med $R.\text{Bokning}$ är att anpassa en multipel regression som beskriver Beläggning. Ekvationen för denna visas i formel 31.

$$Beläggning = \beta_0 + \beta_1 R.\text{FörraÅret} + \beta_2 R.\text{Bokningsläge} + \beta_3 \text{Event} + \beta_4 \text{Högtider} + \beta_5 \text{EH} + n_t$$

(31)

Denna regression innehåller alla förklarande variabler och får därmed saknade värden för december varje år. Den justerade förklaringsgraden för regressionen är $0,7407$ vilket innebär att de förklarande variablerna klarar av att förklarar $74,07\%$ av variationen i responsvariabeln Beläggning. Detta ger en första indikation om att variablerna är lämpliga som inputvariabler till Beläggning.

Utifrån modellen i formel 31 studeras residualerna ($n_t$) och en lämplig modell bestäms. Residualerna $n_t$ för regressionen undersöks och anpassas till en ARMA-modell. SAC och SPAC för residualerna visas i figur 10.
Figur 10 visar att SAC dör ut och SPAC har spik vid tidsförskjutning 1 vilket tyder på att \( n_t \) följer en AR(1) struktur. Då Beläggning uppvisar veckovariation, se bilaga 3, studeras även var sjunde tidsförskjutning, med start på tidsförskjutning 7. SPAC har spik vid säsong 1 och SAC dör ut vilket tyder på en SAR(1). Vidare modellering utgår därför från en SARMA(1, 0)(1, 0)

Den valda SARMA-modellen för \( n_t \) anpassas nu tillsammans med den multipla regressionen till en dynamisk regression och parametrarna skattas samtidigt. När SAC och SPAC ritas upp för den valda SARMA-modellen syns det fortfarande tydlig autokorrelation. På grund av detta undersöks ett flertal olika SARMA-modeller, där den mest lämpliga anses vara en SARMA(3, 0)(1, 0). Ett Ljung-Box-test utförs för den valda modellen för att undersöka om \( a_t \) till den dynamiska regressionen inte kan anses vara vitt brus, det vill säga om \( a_t \) fortfarande har autokorrelation. Hypoteserna för testet visas nedan och resultatet samt autokorrelationerna på tidsförskjutning 1 - 12 visas i tabell 6.

\[ H_0: \text{Residualerna har ingen autokorrelation.} \]
\[ H_a: \text{Residualerna har autokorrelation.} \]

Tabell 6: Ljung-Box-test för autokorrelation för dynamiska regressioner med olika modeller för residualerna.

<table>
<thead>
<tr>
<th>Till tidsförskjutning: DF</th>
<th>P-värde</th>
<th>Autokorrelation</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>2</td>
<td>-0,001</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-0,008</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0,007</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-0,048</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0,087</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0,012</td>
</tr>
<tr>
<td>12</td>
<td>8</td>
<td>0,0069</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0,002</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-0,01</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-0,003</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0,058</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0,061</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-0,043</td>
</tr>
</tbody>
</table>

P-värdena i tabell 6 är mindre än 0,05 därmed kan \( H_0 \) förkastas. Detta indikerar med 95 % säkerhet att residualerna fortfarande är autokorrelade för den dynamiska regressionen med SARMA(3, 0)(1, 0) för \( n_t \). Autokorrelationen verkar vara högre på tidsförskjutningarna 4, 5, 10, 11 och 12 som är tidsförskjutningar vi inte finner en logisk förklaring till jämfört med övriga tidsförskjutningar. Modellen undersöks noggrannare där SAC och SPAC för denna modell visas i figur 11.
Figur 11: SAC och SPAC för modell 1 där residualerna är anpassade med en SARMA(3, 0)(1, 0).

SAC och SPAC i figur 11 tyder på att det finns någon form av autokorrelation kvar vid tidsförskjutning 5 och möjligtvis även i 10 och 15. Då vi inte vet vad detta beror på och då autokorrelationerna ansågs låga för tidsförskjutningar som vi vet vad det är så får detta vara kvar för att inte riskera en överanpassning.

Den slutliga modellen för $n_t$ väljs därför till en SARMA(3, 0)(1, 0) och ekvationen för denna visas i formel 32.

$$\begin{align*}
(1 - \phi_1B - \phi_2B^2 - \phi_3B^3)(1 - \phi_{1,7}B^7)n_t &= \delta + a_t
\end{align*} \tag{32}$$

Kravet på det vita bruset, $a_t$, är att det ska vara oberoende och normalfördelat. Då det nu har fastställts att $a_t$ är godtyckligt oberoende undersöks huruvida normalfördelningsantaget är uppfyllt i figur 12.

Figur 12: Överblick för residualerna för modell 1.
Histogrammet i figur 12 tyder på att \( a_t \) är symmetrisk, men att fördelningen har ganska långa svansar. Den normalfördelade sannolikhetskurvan tyder på att \( a_t \) följer normalfördelningen bra för de mittersta observationerna, men mycket sämre för svansarna. När den normalfördelade sannolikhetskurvan ser ut på detta sätt tyder det på långa svansar. Både histogrammet och sannolikhetskurvan visar alltså på samma typ av problem. Eftersom inget annat problem än detta utläses anses det vara försumbart, och \( a_t \) anses uppfylla normalfördelningsantagandet. Utifrån detta anses modellen välanpassad.

Tabell 7: Skattningar för den dynamiska regressionen, modell 1.

<table>
<thead>
<tr>
<th>Koefficient</th>
<th>Skattning</th>
<th>P-värde</th>
<th>VIF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0,6996</td>
<td>&lt;0,0001</td>
<td></td>
</tr>
<tr>
<td>AR(1)</td>
<td>0,5796</td>
<td>&lt;0,0001</td>
<td></td>
</tr>
<tr>
<td>AR(2)</td>
<td>-0,1292</td>
<td>0,0005</td>
<td></td>
</tr>
<tr>
<td>AR(3)</td>
<td>0,0917</td>
<td>0,004</td>
<td></td>
</tr>
<tr>
<td>SAR(1)</td>
<td>0,1748</td>
<td>&lt;0,0001</td>
<td></td>
</tr>
<tr>
<td>R.FörraÅret</td>
<td>0,7596</td>
<td>&lt;0,0001</td>
<td>1,078</td>
</tr>
<tr>
<td>R.Bokning</td>
<td>0,0001</td>
<td>&lt;0,0001</td>
<td>1,076</td>
</tr>
<tr>
<td>Event</td>
<td>0,0404</td>
<td>&lt;0,0001</td>
<td>1,034</td>
</tr>
<tr>
<td>Högtider</td>
<td>-0,0975</td>
<td>&lt;0,0001</td>
<td>1,027</td>
</tr>
<tr>
<td>EH</td>
<td>0,0485</td>
<td>0,0323</td>
<td>1,009</td>
</tr>
</tbody>
</table>

Tabell 7 visar att alla variabler uppgir signifikans på 5 % signifikansnivå. Att den generella effekten för Högtider är negativ och den generella effekten för Event är positiv visar att variablen EH är viktig för att inte Högtider och Event ska bli missvisande. EH har en effekt större än Event vilket skulle innebära att effekterna för Högtider skulle bli mindre om EH inkluderades i dessa. Variablerna FörraÅret, R.Bokning har en positiv effekt på Beläggning.

VIF-värdena tyder inte på någon multikollinjäritetsproblem för modellen.

Den slutliga modellen presenteras i ekvationsform i formel 33 och 34.

\[
Beläggning = 0,6996 + 0,7596R. FörraÅret + 0,0001R. Bokning + 0,0404Event - 0,0975Högtid + 0,0485EH + n_t \tag{33}
\]

\[
(1 - 0,5796B + 0,1292B^2 - 0,0917B^3)(1 - 0,1748B^2)n_t = a_t \tag{34}
\]

4.1.2 Modell 2

Det som skiljer denna regression från den tidigare i formel 31 är att den inte innehåller R.Bokningsläge, se formel 35.

\[
Beläggning = \beta_0 + \beta_1R. FörraÅret + \beta_2Event + \beta_3Högtider + \beta_4EH + n_t \tag{35}
\]

Denna regression består alltså av värden från och med år 2009 istället för år 2013 som den tidigare modellen samt har värden för december månad. Resultatet från denna modell kommer att imputeras där det saknas värden för modellen med R.Bokning.  Denna modell används alltså för att prognostisera månaden december. Den justerade förklaringsgraden för regressionen i formell 35 är 0,6843 vilket innebär att de förklarande variablerna klarar av att förklara 68,43 % av variationen i responsvariabeln Beläggning. Det visar att regressionen i formel 31 klarar av att förklara mer variation vilket tyder på att variabeln R.Bokning är lämplig att använda för prognos över januari till november även då den innehåller många saknade värden.
Residualerna $n_t$ för regressionen undersöks och en lämplig ARMA-modell anpassas. SAC och SPAC för $n_t$ visas i figur 13.

**Diagram på residualernas korrelation**

![Diagram](image)

**Figur 13:** SAC och SPAC för regressionen i ekvation 35.

Sett till icke-säsong visar figur 13 att SAC dör ut och SPAC har spikar vid tidsförskjutning 1 och möjligtvis 2. För säsongen dör SAC ut och SPAC har en spik på tidsförskjutning 7. Detta tyder på en AR(2) och en SAR(1) med säsongen 7. Vidare modellering utgår från en SARMA(2, 0)(1, 0).7

Den valda modellen för residualerna anpassas nu tillsammans med regressionen i formel 35 till en dynamisk regression och parametrarna skattas samtidigt. Genom att anpassa olika SARMA-modeller fastslogs det att en SARMA(3, 0)(1, 0)7 var bäst lämpad för $n_t$. Liksom tidigare undersöks om det går att fastställa att $a_t$, till den dynamiska regressionen med den valda modellen till $n_t$ (SARMA(3, 0)(1, 0)), är autokorrelerade genom ett Ljung-Box-test för den valda modellen. Testresultat samt autokorrelationerna på tidsförskjutning 1 - 12 visas i tabell 8.

$H_0$: Residualerna har ingen autokorrelation.

$H_a$: Residualerna har autokorrelation.

**Tabell 8:** Ljung-Box-test för autokorrelation för dynamiska regressioner med olika modeller för residualerna.

<table>
<thead>
<tr>
<th>Till tidsförskjutning:</th>
<th>DF</th>
<th>P-värde</th>
<th>Autokorrelation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>&lt;0,001</td>
<td>-0,004 0,00 -0,01 -0,026 0,124 -0,036</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>&lt;0,0001</td>
<td>-0,012 -0,044 0,022 0,06 0,058 -0,016</td>
</tr>
</tbody>
</table>

Inget av p-värdena i tabell 8 är större än 0,05 vilket innebär att $H_0$ kan förkastas med 95 % säkerhet. Detta indikerar med 95 % säkerhet att modells residualer ($a_t$) fortfarande är autokorrelerade.

Autokorrelationen verkar vara hög på framförallt tidsförskjutningarna 5 som är tidsförskjutningar vi inte finner en logisk förklaring till. Då vi genom att testa oss fram inte kunde hitta någon mer lämplig ARMA-modell än SARMA(3, 0)(1, 0), valde vi att gå vidare med denna. I figur 14 visas SAC och SPAC för den dynamiska regressionen med den valda ARMA-modellen.
Figur 14: SAC och SPAC för modell 2 där residualerna är anpassade med en SARMA(3, 0)(1, 0).

SAC och SPAC i figur 14 tyder på att modellen är välanpassad men att det på tidsförskjutning 5 och 10 finns kvar en viss autokorrelation. Då vi inte har någon information om vad detta beror på så anses modellen bra utifrån det som går att ta hänsyn till. För att inte riskera en överanpassad modell så väljs denna modell. Då denna modell endast används för att prognostisera månaden december så anses inte den kvarvarande autokorrelationen så allvarlig.

Den slutliga modellen för $n_t$ är en SARMA(3, 0)(1, 0), och ekvationen för denna visas i formel 36.

$$(1 - \phi_1 B - \phi_2 B^2 - \phi_3 B^3)(1 - \phi_{1,7} B^7)n_t = \delta + a_t$$

(36)

Då $a_t$ godtyckligt kan anses var oberoende efter att ha studerat figur 14 är det även av intresse att se om $a_t$ har normalfördelning. Detta undersöks i figur 15.

Diagnos på residualernas normalfördelnings antagande för modell 2

Figur 15: Överblick för residualerna till modell 2.
Histogrammet i figur 15 tyder på att \( a_t \) följer samma mönster som för modell 1 i avsnitt 4.1.2, det vill säga att fördelningen är symmetrisk men har långa svansar. Även den normalfördelade sannolikhetskurvan tyder på att det största problemet är de långa svansarna. Eftersom de långa svansarna inte är så allvarligt kan \( a_t \) anses uppfylla normalfördelningsantagandet och modellen är därmed välanpassad.

Tabell 9: Skattningar för den dynamiska regressionen, modell 2.

<table>
<thead>
<tr>
<th>Koefficient</th>
<th>Skattning</th>
<th>P-värde</th>
<th>VIF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0,7200</td>
<td>&lt;0,0001</td>
<td></td>
</tr>
<tr>
<td>AR(1)</td>
<td>0,6666</td>
<td>&lt;0,0001</td>
<td></td>
</tr>
<tr>
<td>AR(2)</td>
<td>-0,1669</td>
<td>&lt;0,0001</td>
<td></td>
</tr>
<tr>
<td>AR(3)</td>
<td>0,0807</td>
<td>&lt;0,0001</td>
<td></td>
</tr>
<tr>
<td>SAR(1)</td>
<td>0,2029</td>
<td>&lt;0,0001</td>
<td></td>
</tr>
<tr>
<td>R.FörraÅret</td>
<td>0,8169</td>
<td>&lt;0,0001</td>
<td>1,017</td>
</tr>
<tr>
<td>Event</td>
<td>0,0288</td>
<td>&lt;0,0001</td>
<td>1,025</td>
</tr>
<tr>
<td>Högtider</td>
<td>-0,0845</td>
<td>&lt;0,0001</td>
<td>1,019</td>
</tr>
<tr>
<td>EH</td>
<td>0,0252</td>
<td>0,1806</td>
<td>1,003</td>
</tr>
</tbody>
</table>


Den slutliga modellen presenteras i ekvationsform i formel 37 och 38.

\[
\text{Beläggning} = 0,7200 + 0,8169\text{R.FörraÅret} + \\
+ 0,0288\text{Event} - 0,0845\text{Högtider} + 0,0252\text{EH} + n_t
\]

\[
(1 - 0,6666B + 0,1669B^2 - 0,0807B^3)(1 - 0,2029B^7)n_t = a_t
\]

4.2 Modell rumsintäkter

Tabell 10: Sammanfattning variabler tillhörande modellerna för rumsintäkter.

<table>
<thead>
<tr>
<th>Variablename</th>
<th>Förklaring</th>
<th>Status år 2016</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rumsintäkter</td>
<td>Intäkterna för de sålda rummen för Stockholm som helhet.</td>
<td>Prognostiseras</td>
</tr>
<tr>
<td>SåldaRum</td>
<td>Antalet sålda rum.</td>
<td>Beräknas</td>
</tr>
<tr>
<td>R.Bokning</td>
<td>Bokning justerat för oönskad variation. Skapad.</td>
<td>Känd</td>
</tr>
<tr>
<td>Helg</td>
<td>Indikatorvariabel för dagarna fredag - söndag.</td>
<td>Känd</td>
</tr>
<tr>
<td>Juli</td>
<td>Indikatorvariabel för månaden juli.</td>
<td>Känd</td>
</tr>
<tr>
<td>Event</td>
<td>Evenemang som inte är högtider, t.ex. konsorter eller mässor.</td>
<td>Känd</td>
</tr>
<tr>
<td>Högtider</td>
<td>Evenemang som är högtider, t.ex. påsk och jul.</td>
<td>Känd</td>
</tr>
<tr>
<td>EH</td>
<td>Dagar där Event och Högtider inträffar samtidigt.</td>
<td>Känd</td>
</tr>
<tr>
<td>SåldaRum*Helg</td>
<td>Interaktionsterm som förklarar beroendet i figur 6.</td>
<td>Beräknas</td>
</tr>
</tbody>
</table>

4.2.1 Modell 3

Först anpassas den multipla regressionen för rumsintäkter med R.Bokning och de sju övriga förklarande variablerna. Ekvationen för denna regression visas i formel 39.

\[
\text{Rumsintäkter} = \beta_0 + \beta_1 \text{SåldaRum} + \beta_2 \text{Helg} + \beta_3 \text{Juli} + \beta_4 R.\text{Bokning} + \beta_5 \text{Event} + \\
\beta_6 \text{Högtider} + \beta_7 \text{EH} + \beta_8 \text{SåldaRum} \times \text{Helg} + \epsilon
\] (39)

Då denna regression innehåller R.Bokning kommer den att få saknade värden för åren innan 2013 samt för december varje år. Den justerade förklaringsgraden för regressionen är 0,8939 vilket innebär att de förklarande variablerna klarar av att förklara 89,39 % av variationen i responsvariabeln Rumsintäkter. Sett till detta anses variablerna lämpliga som förklarande varibler till Rumsintäkter.

Utifrån modellen i formel 39 studeras residualerna \(\epsilon\) och en lämplig ARMA-modell till \(\epsilon\) bestäms. SAC och SPAC för residualerna visas i figur 16.

![Diagram](image)

Figur 16: SAC och SPAC för regression i formel 39.

SAC och SPAC i figur 16 tyder på att det finns autokorrelation i \(\epsilon\). Även Rumsintäkter tycks följa veckovariationen, se bilaga 5, och det finns autokorrelation på var sjunde tidsförskjutning. På icke-säsongstidsförskjutningar avtar SAC och SPAC har spik vid tidsförskjutning 1. Detta pekar på att en AR(1) är lämplig. Vid säsongstidsförskjutningarna tycks SAC avta och SPAC har spik vid säsongstidsförskjutning 1 vilket tyder på en SAR(1). Utifrån detta så anpassas en dynamisk regression med en SARMA(1, 0)(1, 0) för residualerna.
SARMA-modellen anpassas tillsammans med regressionen i formel 39 och parametrarna skattas samtidigt i form av en dynamisk regression. Förutom SARMA(1, 0)(1, 0), undersöktes ytterligare SARMA-modeller, och en SARMA(2, 0)(3, 0) anses mest lämpad för $n_t$. För att undersöka om $a_t$, i den slutliga dynamiska regressionen med den valda SARMA-modellen, kan anses vara autokorrerade utförs ett Ljung-Box-test. Autokorrelationerna för tidsförskjutning 1-12 samt resultatet för testet visas i tabell 11.

$$H_0: \text{Residualerna har ingen autokorrelation.}$$

$$H_a: \text{Residualerna har autokorrelation.}$$

Tabell 11: Tabell över p-värden för Ljung-Box-test.

<table>
<thead>
<tr>
<th>Tidsförskjutning</th>
<th>DF</th>
<th>P-värde</th>
<th>Autokorrelation</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>1</td>
<td>0,0472</td>
<td>0,007</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-0,027</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0,022</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0,002</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-0,031</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0,036</td>
</tr>
</tbody>
</table>

P-värden i tabell 11 till tidsförskjutning 12 för modellen SARMA(2, 0)(3, 0) är större än 0,05 vilket innebär att $H_0$ inte kan förkastas. Detta indikerar att det inte går att säkerställa med 95% säkerhet att $a_t$ är autokorrelerade. Autokorrelationerna anses relativt låga på alla tidsförskjutningar.

Utifrån detta så anpassas en dynamisk regression med modellen SARMA(2, 0)(3, 0) för residualerna.

**Diagnos på residualernas korrelationer för modell 3**

SAC och SPAC för denna modell visas i figur 17.

Diagnos på residualernas korrelationer för modell 3

Residualerna, i figur 17, för den slutliga modellen tyder på att modellen är välanpassad och det verkar inte återstå någon betydande autokorrelation. Den slutliga modellen för residualerna är en SARMA(2, 0)(3, 0), och ekvationen för denna syns i formel 40.

$$n_t = \delta + a_t$$

Utifrån figur 17 kan $a_t$ anses vara oberoende och det undersöks om $a_t$ även kan anses ha normalfördelning. Detta studeras visuellt i figur 18.
Figur 18: Överblick för residualerna till modell 1 för Rumsintäkter där residualerna är anpassade med en SARMA(2, 0)(3, 0).

Histogrammet i figur 18 tyder på att $a_t$ är relativt normalfördelad, men att den återigen har långa svansar. Den normalfördelade sannolikhetskurvan visar också på detta, men det verkar främst vara ett fåtal extrempunkter som gör att sannolikhetskurvan avviker. I övrigt ser fördelningen ut att vara symmetrisk. Dessa grafer tyder på att $a_t$ har en godtycklig normalfördelning och att modellen kan anses uppfylla normalfördelningsantagandet.

Då $a_t$ uppfyller antagandena så anses modellen välanpassad.

Tabell 12: Skattningar för den slutliga dynamiska regressionen, Modell 3.

<table>
<thead>
<tr>
<th>Koefficient</th>
<th>Skattning</th>
<th>P-värde</th>
<th>VIF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-2 599 781</td>
<td>&lt;0,0001</td>
<td></td>
</tr>
<tr>
<td>AR(1)</td>
<td>0,9440</td>
<td>&lt;0,0001</td>
<td></td>
</tr>
<tr>
<td>AR(2)</td>
<td>-0,2056</td>
<td>&lt;0,0001</td>
<td></td>
</tr>
<tr>
<td>SAR(1)</td>
<td>0,3823</td>
<td>&lt;0,0001</td>
<td></td>
</tr>
<tr>
<td>SAR(2)</td>
<td>0,2423</td>
<td>&lt;0,0001</td>
<td></td>
</tr>
<tr>
<td>SAR(3)</td>
<td>0,2512</td>
<td>&lt;0,0001</td>
<td></td>
</tr>
<tr>
<td>SåldaRum</td>
<td>1 458,30</td>
<td>&lt;0,0001</td>
<td>2,927</td>
</tr>
<tr>
<td>Helg</td>
<td>2 455 598</td>
<td>&lt;0,0001</td>
<td>16,651</td>
</tr>
<tr>
<td>juli</td>
<td>-592 510</td>
<td>0,0278</td>
<td>1,023</td>
</tr>
<tr>
<td>R,Bokning</td>
<td>1 112,80</td>
<td>&lt;0,0001</td>
<td>1,182</td>
</tr>
<tr>
<td>Event</td>
<td>659 527</td>
<td>&lt;0,0001</td>
<td>1,092</td>
</tr>
<tr>
<td>Högtid</td>
<td>-539 656</td>
<td>&lt;0,0001</td>
<td>1,083</td>
</tr>
<tr>
<td>EH</td>
<td>-54 153,8</td>
<td>0,7446</td>
<td>1,026</td>
</tr>
<tr>
<td>SåldaRum*Helg</td>
<td>-382 560</td>
<td>&lt;0,0001</td>
<td>15,569</td>
</tr>
</tbody>
</table>

Tabell 12 visar att alla variabler förutom $EH$ är signifikanta på 5 % signifikansnivå. $EH$ har en effekt mitt emellan $Event$ och $Högtider$ vilket skulle innebära att effekterna för $Event$ och $Högtider$ skulle bli mindre om $EH$ inkluderades i dessa. Därmed behålls $EH$ i modellen. Det går vidare att utläsa att variablerna $SåldaRum$, $R.Bokning$ och $Event$ har en positiv effekt på $Rumsintäkter$ och variablerna $Juli$,

VIF-värdena i tabell 12 tyder på att det inte finns multikollinjaritetsproblem i modellen. De höga VIF-värdena för interaktionstermen och Helg beror på att dessa korrelerar med varandra men detta anses inte som ett problem vid fallet med interaktionsterm.

Den slutliga modellen presenteras i ekvationsform i formel 41 och 42.

\[
\text{Rumsintäkt} = -1845439 + 1438,09 \text{SälldaRum} + 1152,86 R.\text{Bokning} + \\
+ 700686 \text{Event} - 545428 \text{Högtid} - 35080,9 \text{EH} - \\
-830678 \text{Juli} + 1984400 \text{Helg} - 333,339 \text{SälldaRum} \times \text{Helg} + n_t
\]

\[
(1 - 0,9600B + 0,2089B^2)(1 - 0,3735B^7 - 0,2205B^{14} - 0,1691B^{21})n_t = \delta + a_t
\]

4.2.2 Modell 4

Denna modell innehåller inte R.Bokning vilket innebär att modellen består av fler värden än tidigare modell, eftersom R.Bokning inte finns innan år 2013. Den multipla regressionens ekvation visas i formel 43.

\[
\text{Rumsintäkt} = \beta_0 + \beta_1 \text{SälldaRum} + \beta_2 \text{Helg} + \beta_3 \text{Juli} + \beta_4 \text{Event} + \\
\beta_5 \text{Högtider} + \beta_6 \text{EH} + +\beta_7 \text{SälldaRum} \times \text{Helg} + n_t
\]

Resultatet från denna modell kommer att imputeras där det saknas värden för modellen med R.Bokning, det vill säga i december för år 2016. Den justerade förklaringsgraden visar att modellen klarar 86,45 % av variationen i responsvariabeln Rumsintäkter. Detta är något lägre jämfört med modellen i formel 39 som inkluderar R.Bokning.

Vidare studeras SAC och SPAC för residualerna från denna multipla regression i figur 19 för att undersöka vilken ARMA-modell som är lämplig.

**Diagnos på residualernas korrelation**

![Figur 19: SAC och SPAC för regressionen i formel 43.](image)

SAC och SPAC i figur 19 tyder på att det finns autokorrelation i \(n_t\), både för säsong och icke-säsong. På icke-säsongen avtar SAC och SPAC har en spik vid tidsförskjutning 1 vilket tyder på en AR(1). Vid säsongstidsförskjutningarna avtar SAC och SPAC har spik vid säsong 1 som pekar på att en SAR(1) bör
anpassas. Utifrån detta så utgår vidare modellering med en dynamisk regression från en SARMA(1, 0)(1, 0), för residualerna.

Den dynamiska modellen anpassas nu genom att parametrarna i den multipla regressionen samt de i SARMA-modellen skattas samtidigt. Förutom det som utlåsades direkt från figur 19 prövades även ytterligare modeller. Dessa var SARMA(2, 0)(1, 0), SARMA(2, 0)(2, 0), och SARMA(2, 0)(3, 0), där den sistnämnda ansågs bäst. För att undersöka om det går att fastställa att \( a_t \) är autokorrelerade så utförs ett Ljung-Box-test vars testresultat visas i tabell 13.

\[
H_0: \text{Residualerna har ingen autokorrelation.} \\
H_a: \text{Residualerna har autokorrelation.}
\]

Tabell 13: Tabell över p-värden för Ljung-Box-test.

<table>
<thead>
<tr>
<th>Till tidsförskjutning:</th>
<th>DF</th>
<th>P-värde</th>
<th>Autokorrelation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>6</td>
<td>&lt;0,0001</td>
<td>-0,011</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>&lt;0,0001</td>
<td>0,004</td>
</tr>
</tbody>
</table>

P-värdet för den dynamiska regressionen i tabell 13 är mindre än 0,05, vilket innebär att \( H_0 \) kan förkastas. Detta indikerar att det går att säkerställa med 95 % säkerhet att \( a_t \) är autokorrelerade. Den mesta autokorrelationen verkar ligga på tidsförskjutningarna 3, 4 och 5 och även kring 10 och 12. Då vi inte funnit en logisk förklaring till dessa tidsförskjutningars korrelation så undersöks den valda dynamiska regressionen ytterligare genom SAC och SPAC i figur 20.

Diagnos på residualernas korrelationer för modell 4

SAC och SPAC i figur 20 tyder på att modellen är välanpassad bortsett från tidsförskjutningarna runt 4 och 5 som vi inte klarar av att förklara. Liksom tidigare får denna autokorrelation kvarstå då det finns risk för att överanpassa modellen. Då denna modell endast används för att prognostisera månaden december så anses inte den kvarvarande autokorrelationen så allvarlig.

Med detta i åtanke anses SARMA(2, 0)(3, 0), vara lämplig för att modellera \( n_t \).

Ekvationen för denna SARMA visas i formel 44.

\[
(1 - \phi_1 B - \phi_2 B^2)(1 - \phi_{1,7} B^7 - \phi_{2,7} B^{14} - \phi_{2,7} B^{21})n_t = \delta + a_t
\]  

(44)

I figur 21 studeras huruvida \( a_t \) kan anses vara normalfördelad.
Figur 21: Överblick för residualerna till modell 2 för Rumsintäkter där residualerna är anpassade med en SARMA(2, 0)(3, 0). Histogrammet visar att $a_t$ har symmetri, men väldigt långa svansar. Även den normalfördelade sannolikhetskurvan tyder på detta. $a_t$ är på gränsen till icke-normalfördelad på grund av de extrema svansarna, men på grund av den tydliga symmetrin anses $a_t$ vara normalfördelad, om än väldigt godtyckligt och modellen uppfyller normalfördelningsantagandet.

Då SAC och SPAC visar att $a_t$ är relativt oberoende och då normalfördelningsantagandet anses uppfyllt så anses modellen relativt välhanpassad.


<table>
<thead>
<tr>
<th>Koefficient</th>
<th>Skattning</th>
<th>P-värde</th>
<th>VIF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-1 685 086</td>
<td>&lt;0,0001</td>
<td></td>
</tr>
<tr>
<td>AR(1)</td>
<td>1 1786</td>
<td>&lt;0,0001</td>
<td></td>
</tr>
<tr>
<td>AR(2)</td>
<td>-0,3795</td>
<td>&lt;0,0001</td>
<td></td>
</tr>
<tr>
<td>SAR(1)</td>
<td>0,3025</td>
<td>&lt;0,0001</td>
<td></td>
</tr>
<tr>
<td>SAR(2)</td>
<td>0,2183</td>
<td>&lt;0,0001</td>
<td></td>
</tr>
<tr>
<td>SAR(3)</td>
<td>0,2256</td>
<td>&lt;0,0001</td>
<td></td>
</tr>
<tr>
<td>SälldaRum</td>
<td>1 459,90</td>
<td>&lt;0,0001</td>
<td>2,278</td>
</tr>
<tr>
<td>Helg</td>
<td>1 859 066</td>
<td>&lt;0,0001</td>
<td>10,254</td>
</tr>
<tr>
<td>juli</td>
<td>-340 813</td>
<td>0,0523</td>
<td>1,016</td>
</tr>
<tr>
<td>Event</td>
<td>303 877</td>
<td>&lt;0,0001</td>
<td>1,121</td>
</tr>
<tr>
<td>Högtid</td>
<td>-117 589</td>
<td>0,1113</td>
<td>1,042</td>
</tr>
<tr>
<td>EH</td>
<td>-118 487</td>
<td>0,3412</td>
<td>1,013</td>
</tr>
<tr>
<td>SälldaRum*Helg</td>
<td>-362,560</td>
<td>&lt;0,0001</td>
<td>9,860</td>
</tr>
</tbody>
</table>

Alla variabler i tabell 14 är signifikanta förutom EH och Högtid. Då vi tidigare sett att dessa variabler har en effekt på Rumsintäkter så behålls de ändå. VIF-värdena är höga för variablerna Helg och interaktionsstermen SälldaRum*Helg vilket beror på att variabeln Helg finns med i båda variablerna. Vidare utläses det att variablerna SälldaRum och Event har en positiv effekt på Rumsintäkter och

Den slutliga modellen presenteras i ekvationsform i formel 45 och 46.

\[
Rumsintäkt = -1874418 + 1482,13SåldaRum + 307111Event - 98808,6Högtid - 141241EH - 449832Juli + 1811932Helg - 349,840SåldaRum * Helg + n_t
\]

\[
(1 - 1,1762B + 0,3694B^2)(1 - 0,3011B^7 - 0,2157B^{14} - 0,1844B^{21})n_t = \delta + a_t
\]

4.3 Prognoser
I detta avsnitt redovisas prognosen för Beläggning samt Rumsintäkter, Snittpris och RevPAR år 2016 som beräknats utifrån tidigare nämnda prognos.

4.3.1 Beläggning
Prognosen för Beläggning visas nedan i figur 22. Som referens plottas även Beläggning för år 2013-2015. Fem dagar under april till juni fick prognoser som var något högre än ett, vilket skulle innebära att det skulle sälts fler rum än vad som fanns disponibelt den dagen. Eftersom detta är orimligt har dessa dagar justerats ner manuellt till 1, eftersom Beläggning aldrig kan vara högre än 100 %.


Figur 22 visar att prognosen för första halvan av året följer mönstret för de tre tidigare åren, men för den senare delen av året är prognosen lägre än vad som kan förväntas utifrån en visuell analys. Detta kan bero på att bokningsläget inte justerats upp tillräckligt för de senare månaderna och att det är detta som återspeglas här.
Prognosen för 2016 grupperas på eventtyperna för att visuellt undersöka om det finns något mönster att utläsa. Detta visas i figur 23, där de gråa punkterna i figuren indikerar dagar då ingen typ av evenemang inträffar.

![Diagram: Beläggningen uppdelat efter evenemangstyp](image)

Figur 23: Prognosen år 2016 för Beläggning uppdelat efter evenemangstyp.


Figur 23 tyder på en stark positiv trend från mitten av januari och har en topp i mitten av februari. Under andra halvan av februari minskar Beläggning till en lägre nivå som varar fram till mitten av mars. I mitten av mars ökar Beläggning till en högre nivå som varar fram till mitten av april och Beläggning går sedan ned igen liknande som i mitten av februari. I början av maj tilltar en stark positiv trend och Beläggning tycks öka fram till början på juni. I slutet av juni startar en svagt negativ trend som varar året ut med en ökad varians för varje månad.

4.3.2 Rumsintäkter

Prognosen i figur 24 visar att Rumsintäkter följer det generella årsmönstret, men skattar intäkterna för hösten och vintern något lägre än tidigare år. Modellen har inte heller klarat av att skatta de högsta topparna.

Rumsintäkter för år 2016 visas grupperade efter evenemangstypen i figur 25. Liksom tidigare innebär de gråa punkterna att inget evenemang inträffar den dagen.


4.3.3 Snittpris
Snittpris beräknas utifrån Rumsintäkter enligt formel 2 och prognosen tillsammans med värden för tidigare år visas i figur 26.

![Snittpriset med prognos för 2016](image.jpg)


Prognosen för Snittpris i figur 26 visar att Snittpris år 2016 generellt följer samma årliga mönster som tidigare år men har en svagare negativ trend i slutet av året jämfört med tidigare år. De prognostiserade värdena är lägre under det första halvåret jämfört med tidigare år.

Prognosen för Snittpris i figur 27 verkar, i början av året, påverkas positivt av eventen jämfört med sista halvåret då Snittpris varierar oberoende av Event. Högtider verkar ha en negativ effekt på Snittpris då alla fem högtider som är registrerade för 2016 ligger vid den lägre nivån i den aktuella trenden.


4.3.4 RevPAR

För att få fram RevPAR divideras Rumsintäkter med Disponibla rum, se formel 3. I figur 28 illustreras prognosen för RevPAR.


Figur 29: Prognosen år 2016 för RevPAR.
Enligt figur 29 verkar Event ge ett högt RevPAR i förhållande till månaden och Högtider verkar minska RevPAR i förhållande till månaden. Det är svårt att utläsa någon generell effekt från EH då det endast finns tre observationer för en sådan händelse och dessa visar inte på samma effekt. Ingen av observationerna för EH verkar heller bryta den aktuella trenden.

5 Diskussion

I detta kapitel diskuteras resultatets innebörd och metodernas lämplighets utvärderas. Då det finns fyra slutliga modeller, två stycken var för Beläggning respektive Rumsintäkter, presenteras nedan tabell 15 som visar benämningen för de olika modellerna i detta kapitel.

Tabell 15: Benämningar för uppsatsens modeller.

<table>
<thead>
<tr>
<th>Benämning</th>
<th>Modell</th>
</tr>
</thead>
<tbody>
<tr>
<td>B1</td>
<td>Beläggning med R.Bokning (Modell 1)</td>
</tr>
<tr>
<td>B2</td>
<td>Beläggning utan R.Bokning (Modell 2)</td>
</tr>
<tr>
<td>R1</td>
<td>Rumsintäkter med R.Bokning (Modell 3)</td>
</tr>
<tr>
<td>R2</td>
<td>Rumsintäkter utan R.Bokning (Modell 4)</td>
</tr>
</tbody>
</table>

5.1 Datamaterialets lämplighet

I allra största utsträckning har datamaterialet och dess variabler varit väl lämpad för att skapa modeller och göra prognoser av nyckeltalen, men det finns vissa saker som kan förändras för att ytterligare förbättra prognoserna.


Vidare har vi sett att evenemangen ibland skrivits in utan marknad eller med felaktigt datum, då till exempel ett evenemang pågick under ett år och en dag. Det är en klar minoritet av evenemangen som haft felaktig information, men det är något som borde prioriteras för få så tillförlitliga skattningar som möjligt.


5.2 Modellernas lämplighet
I detta avsnitt diskuteras varför just dynamisk regression valdes samt tankar omkring modellerna B1, B2, R1 och R2 och vad som kan förbättras med dessa.

5.2.1 Dynamisk regression


Innan vi släppte idén med TF-modellen var tanken att vår modell skulle vara en sammanslagning av en TF-modell och en interventionsmodell. Interventionsmodeller används dock oftast för att markera en eller ett fåtal händelser som har stort inflytande på tidsserien, som till exempel strejker eller ändrad företagspolicy. Vi ville dock mata en generell effekt för fler än 500 event, vilket inte passade interventionsmodellen som mäter effekten för individuella eller direkt efterföljande händelser. Även i detta avseende passade den dynamiska regression bättre för vårt ändamål.

5.2.2 Beläggning
Enligt SAC och SPAC i figur 11 för B1 och figur 14 för B2 har residualerna för båda modellerna autokorrelation. Detta antas tyda på att det finns ytterligare säsongsmönster som modellerna inte lyckats förklara, då det finns signifikant autokorrelation var femte tidsförskjutning och de variabler vi har haft till förfogande har vi inte lyckats utskilja någon variabel som kan förklara detta säsongsmönster. För att ta reda på vad detta beror på bör en person med djupare insikt inom hotelmarknaden rådfrågas för att undersöka om det finns någon logisk förklaring till mönstret, då våra vidare analyser på detta inte gett något resultat. Dock har vi lyckats få bort autokorrelationen på de tidsförskjutningarna som kunde utläses innan den dynamiska regressionen anpassades. Förutom detta kan det vita bruset anses vara approximativt normalfördelat vilket pekar på att modellerna är lämpliga för datamaterialet.

Det finns dock utrymme för att förbättra modellerna för Beläggning, och vi tror att det främsta förbättringsområdet ligger på att bearbeta och eventuellt hitta nya variabler som kan förklara denna autokorrelation snarare än att använda en annan metodik för att modellera sambanden.

5.2.3 Rumsintäkter
Även för de slutliga modellerna R1 och R2 uppvisar SAC och SPAC i figur 17 och 20 ett säsongsmönster med säsongslängd fyra eller fem. Anledningen till att den kvarvarande autokorrelationens säsongslängd inte är exakt samma som för B1 och B2 kan bero på att Rumsintäkter även tar hänsyn till Helg och Juli, som möjligtvis har något samband med en latent variabel som påverkar det som ger
säsongslängden fem. Även här bör en person kunnig inom hotellmarknaden rådfrågas för att ta reda på vilken ytterligare variabel som skulle kunna förklara detta.

Ljung-Box-testet blir icke-signifikant för R1 men signifikant för R2, vilket tyder på att $R.Bokning$ klarar av att förklara en del autokorreleras för $Rumsintäkter$. När det vita bruset plottas mot de anpassade värdena har R2 i figur 21 en mer skev fördelning än R1 i figur 18, vilket tyder på att $R.Bokning$ gör det vita bruset mer stationärt. Då den kvarvarande autokorrelationen inte kan förklaras med hjälp av de variabler som vi haft tillhanda anses det vita bruset vara godtyckligt oberoende samt relativt normalfördelade. Med detta i åtanke betraktas modellerna som välanpassade.

Liksom för modellerna för Beläggning finns möjlighet till förbättring, och vi tror som tidigare att det sker främst genom att hitta ytterligare förklarande variabler som hanterar autokorrelationen som finns på säsongstidsförskjutning 4 och 5.

5.3 Resultat

Här tolkas resultatet för modellerna och prognoserna mer djupgående än tidigare samt innebörden av resultatet diskuteras.

5.3.1 Modell för beläggning

Utskrifterna för de dynamiska regressionernas koefficienter (tabell 7 för B1 och tabell 9 för B2) visar att $R.Förråret$ har en positiv effekt på Beläggning. Vi har också sett att Beläggning har en månadsvariation, och genom att ta hänsyn till $R.Förråret$ som är en tidsjusterad variabel av Beläggning borde denna variabel klara av att i alla fall fånga upp en viss del av det årsmönstret.


Effekten för $R.Bokning$ är väldigt svag men positiv vilket tyder på att bokningsläget möjligtvis kan indikera hur den generella nivån för nästkommande år kommer att se ut.

5.3.2 Modell för rumsintäkter


men att priset inte spelar samma roll under vardagar då det oftast är affärsresande och dylikt som
bor på hotell under vardagar och således inte tar samma hänsyn till priset då de bokar hotellrum.

Skattningarna för evenemangvariablerna indikerar att Rumsintäkter tycks öka vid Event, minas vid
Högtider och även minska då det sker ett event samtidigt som en högtid. Även dessa skattningar
känns logiska då man kan tänka sig att det kommer folk från andra städer då det sker event och
många väljer möjligtvis att övernatta. Att Rumsintäkter sjunker på högtiderna kan förklaras genom
att folk kanske tenderar att fira stora högtider så som påsk och jul hemma med familjen och därmed
inte bor på hotell. Att efterfrågan då sjunker leder till att hotellen måste sänka priserna vilket i sin tur
innebär att Rumsintäkter minskar.

Den negativa effekten för variabeln Juli visar att Rumsintäkter minskar för denna månad även då det
sälj många rum den månaden som även visas i figur 7. Detta kan möjligtvis bero på att juli är en
vanlig semestermånad då folk kan tänkas åka utomlands eller ut på landet. Detta leder till att
storstaden bli mindre attraktiv och priserna måste således justeras efter efterfrågan.

5.3.3 Granskning av prognoser

Prognoserna för Beläggning och RevPAR verkar följa tidigare årsmönster relativt bra men med en
lägre variation framförallt i slutet av året. Detta var ganska väntat då data endast finns för åren 2013-
2015 för modellerna med R.Bokning och dessa modeller används för att prognostisera större delen
av året (jan-nov). Vi har även upptäckt felaktigheter i data för Event och Högtider där det helt enkelt
missats att lägga in evenemang. Detta medför att skattningarna inte blir lika bra och därmed inte
heller prognosen. Vi har valt att inte ändra i data då vi ansåg att man då borde se över hela
datamaterialet vilket vi inte fann tid till att göra.

Då vi först prognostiserar SälldaRum för att sedan använda detta till att prognostisera Rumsintäkter
som används för att beräkna Snittpris medför att den inte blir lika tillförlitlig som prognoserna för
Beläggning och RevPAR. Utifrån detta så var det väntat att årsmönstret inte skulle fångas upp i
samma utsträckning och att variationen skiljer sig en del från tidigare år.

Som tidigare nämnt så förväntas prognoserna bli bättre för varje år då mer data kommer finnas
6 Slutsatser

Här presenteras svaren på de fyra frågeställningar som uppsatsen undersökt.

Hur ser de slutliga modellerna ut för Beläggning och Rumsintäkter?

Modellen med \textit{R.Bokning} som prognostiserar \textit{Beläggning} för januari till november innehåller de förklarande variablerna \textit{R.Förråret}, \textit{R.Bokning}, \textit{Event}, \textit{Högtider} och \textit{EH} där feltermerna följer en \textit{AR}(3)-struktur för icke säsong och en säsongs-struktur i form av en \textit{AR}(1) med en säsongslängd av \(7\).

Modellen utan \textit{R.Bokning} som prognostiserar \textit{Beläggning} för månaden december innehåller de förklarande variablerna \textit{R.Förråret}, \textit{Event}, \textit{Högtider} och \textit{EH} där feltermerna följer en \textit{AR}(3)-struktur för icke säsong och en säsongs-struktur i form av en \textit{AR}(1) med en säsongslängd av \(7\).


Hur ser beläggningen ser på dagsnivå för hotellbranschen i Stockholm år 2016?

Beläggningen för 2016 tycks följa tidigare års mönster genom hela året. Den verkar gå ner i början av januari för att sedan ha en positiv trend från mitten av januari och toppa i slutet av månaden. I mitten av februari tycks beläggningen minska till en lägre nivå som håller i sig fram till mitten av mars. I mitten av mars utläses sedan en ny topp som håller i sig mars ut. Första halvan av april verkar ha en lägre beläggning för att sedan följas upp av en positiv trend som tilltar från mitten av april som, med mindre stagneringar, håller i sig fram till början av maj månad. Från slutet an juni månad tilltar en svag negativ trend som håller i sig året ut med en ökad varians för varje månad.

Hur ser snittpriset per sålt rum ut på dagsnivå för hotellbranschen i Stockholm år 2016?

Snittpriset för 2016 tycks följa tidigare års mönster genom hela året. Den verkar gå ner i början av januari för att sedan ha en positiv trend från mitten av januari och toppa i mitten av februari. Efter toppen i mitten februari tycks snittpriset minska radikalt till en nivå som håller i sig fram till början av mars. Snittpriset tilltar sedan och har en topp i slutet av månaden. Snittpriset tycks sedan avta och gå ner till samma nivåer i början av april. I slutet av april månad tycks en ny positiv trend tillta som håller i sig fram till början av juni månad. I slutet av juni hoppar snittpriset ned till ett lågt snittpris som varar under hela juli månad för att sedan hoppa upp till en genomsnittlig nivå i början av augusti. Från augusti start tycks snittpriset variera konstant kring ca 1200 kr.

Vad är intäkten per disponibelt rum på dagsnivå för hotellbranschen i Stockholm år 2016?

7 Referenslista


Bilaga 1
Bilaga 2

Bilaga 3
Beläggningen per veckodag.
Bilaga 4
SåldaRum mot beläggnings uppdelen efter andel besökare från Stockholm.
Bilaga 5
Rumsintäkterna grupperat på veckodag.

![Boxplot of rooms rental per day](image-url)