

Kandidatuppsats i Statistik

Statistisk tidsserieanalys av skillnader i partikelhalter mellan en korsning och en väglänk i Stockholm

En modellering av felterm med GARCH och Cochrane-Orcutt

Eric Herwin
Albin Västerlund



Avdelningen för Statistik och Maskininlärning
Institutionen för datavetenskap
Linköpings universitet

Vårterminen, 2017 | LIU-IDA/STAT-G-17/002-SE

Handledare: Bertil Wegmann

Examinator: Ann-Charlotte Hallberg

Abstract

In Sweden, the assumption is that there is no difference in particle halt between an intersection and a road link and this assumption is mentioned among government institutions. The interest to investigate if statistical differences in particle halt between an intersection and a road link arises when environmental questions concerning intersections has become more interesting.

Data that is used in the thesis is between the 1:th of March 2016 at 13:15 to the 1:th of May 2016 at 00:00 and is measured at three measuring stations that are closely positioned. The stations that measure the particle halt is on the intersection at Hornsgatan 104 and the road link at Hornsgatan 110 where data of the particles with dimensions 1, 2.5 and 10 micrometer in diameter is measured. Meteorological variables is measured at Torkel Knutssongata and the amount of traffic is measured at Hornsgatan 110 and indicates if there is a high or low amount of traffic.

The variables that are used to model the difference are produced by subtracting data from the road link of the data from the intersection. A difference between the intersection and the road link is made and modeling of the difference is possible. One difference is made for each of the particles and the differences has a variance that is not constant and sudden extreme values can appear.

Time serie regressions are used for modeling the three particles (< 1 , 1-2.5 and 2.5-10 micrometer in diameter). Methods such as GARCH and Cochrane-Orcutt are used in the error term where autocorrelation and heteroscedasticity are needed to be taken into account.

For particles with dimensions < 1 and 1-2.5 micrometers in diameter, explained by different time variables, there is a variation of the difference over time and a significant difference between the intersection and the road link exists. For particles with dimensions 2.5-10 micrometer in diameter there is a statistical difference in the month of April. The models with several explanatory variables shows that there is a number of significant variables that can explain the general difference for each particle halt. Those variables are relative humidity, maximum wind and the direction of the wind.

Sammanfattning

Bland statliga institutioner i Sverige existerar antagandet om att det inte finns en skillnad i partikelhalt mellan en korsning och en väglänk. Intresset av att undersöka statistiska skillnader i partikelhalten mellan en korsning och en väglänk har uppkommit då miljörelaterade frågor rörande korsningar har blivit mer relevant att studera.

Data i uppsatsen är mellan 1:a mars 2016 klockan 13:15 till och med den 1:a maj 2016 klockan 00:00 och mäts på tre mätstationer som ligger nära varandra. Mätstationerna som mäter partikelhalt finns på korsningen vid Hornsgatan 104 och väglänken vid Hornsgatan 110 där data om partiklar med måtten 1, 2.5 respektive 10 mikrometer i diameter samlas in. Meteorologiska variabler mäts på Torkel Knutssonsgata och mängden trafik mäts på Hornsgatan 110 och indikerar på hög eller låg trafik.

Variablerna som används för att modellera skillnaderna i uppsatsen är framtagna genom att subtrahera data från väglänken av data från korsningen. En differens mellan korsningen och väglänken skapas och modellering av skillnaden blir möjlig. En differens skapas för varje partikelmått och differenserna har en varians som inte är konstant och oväntade extremvärden förekommer.

Modellering för de tre partikelmåtten (< 1 , 1-2.5 och 2.5-10 mikrometer i diameter) sker med tidsserieregressioner. Metoder som GARCH och Cochrane-Orcutt används i feltermen där autokorrelation och heteroskedacitet behövs tas hänsyn till.

För partiklar med måtten < 1 och 1-2.5 mikrometer i diameter, som förklaras av olika tidsvariabler, varierar skillnaden över tid och en signifikant skillnad mellan korsningen och väglänken finns. Partiklar med 2.5-10 mikrometer i diameter påvisas att ha en skillnad främst under april månad.

Modellerna med fler variabler visar ett antal signifikanta variabler som kan förklara skillnaden mellan korsningen och väglänken. Gemensamma variabler som påverkar skillnaden för varje partikelhalt är relativ fuktighet, maxvind och vindriktning.

Förord

Uppsatsen är ett examensarbete i kandidatprogrammet för Statistik och dataanalys vid Linköpings universitet. Uppdragsgivaren är Statens väg- och transportforskningsinstitut (VTI) och kontaktpersonen är Sara Janhäll.

Ett stort tack går till VTI som har tillhandahållit oss med data till uppsatsen och Sara Janhäll som har varit vår handledare på VTI och har varit tillgänglig via mejl samt möten. Vi vill tacka vår handlare Bertil Wegmann på Linköpings Universitet som under uppsatsens gång har väglett och hjälpt oss genom olika problem som har uppkommit. Vi vill också tacka våra opponenter Sara Jesperson och Sara Johansson för att ha kommit med förslag för förbättringar till uppsatsen.

Innehåll

1	Introduktion	1
1.1	Bakgrund	1
1.2	Tidigare studier	2
1.3	Uppdragsgivare	2
1.4	Syfte	2
1.4.1	Frågeställningar	3
1.5	Etiska och samhälleliga aspekter	3
2	Data	5
2.1	Responsvariabler	5
2.2	Förklaringsvariabler	7
2.3	Imputering	10
3	Metod	13
3.1	Tidsserieregression	13
3.1.1	Modellering för heteroskedasticitet	13
3.1.2	Modellering för autokorrelation	15
3.1.3	Bakåteleminering	16
3.1.4	Partiellt F-test	17
3.2	Implementering i R	17
3.2.1	Paket	17
4	Resultat	19
4.1	Analys av $PM_{1(D)}$	19
4.1.1	Modell med tidsvariabler	19
4.1.2	Modell med förklaringsvariabler	25
4.2	Analys av $PM_{1-2.5(D)}$	28
4.2.1	Modell med tidsvariabler	28
4.2.2	Modell med förklaringsvariabler	34
4.3	Analys av $PM_{2.5-10(D)}$	37
4.3.1	Modell med tidsvariabler	37
4.3.2	Modell med förklaringsvariabler	43
5	Diskussion	47
5.1	Resultatdiskussion	47
5.2	Metoddiskussion	48
6	Slutsats	49
6.1	Finns det någon signifikant skillnad över tid i halter av olika partiklar mellan korsningen och väglänken?	49

6.2	Har meteorologiska, trafikrelaterade och tidsrelaterade variabler någon påverkan på skillnader i partikelhalter mellan korsningen och väglänken? .	49
-----	--	----

Bilaga	52
---------------	-----------

A	Tabeller	i
----------	-----------------	----------

A.1	$PM_{1(D)}$	i
A.2	$PM_{1-2.5(D)}$	ii
A.3	$PM_{2.5-10(D)}$	iii

B	Kod i R	v
----------	----------------	----------

B.1	Modell	v
B.2	ARCH	v
B.3	Cochrane-Orcutt	vi

Figurer

2.1	Responsvariablerna över tid	6
2.2	Medelvärdet av klockslag för varje dag med sinuskurva	9
2.3	Sinuskurva för kl 09-18:00	10
4.1	Residualer över tid av $PM_{1(D)}$ modell	20
4.2	ACF och PACF för fastställande av GARCH för $PM_{1(D)}$	20
4.3	ACF och PACF för fastställande av Cochrane-Orcutt för $PM_{1(D)}$	21
4.4	Residualer över tid korrigerade för autokorrelation och heteroskedasticitet av $PM_{1(D)}$ modell	22
4.5	ACF och PACF korrigerade för autokorrelation och heteroskedasticitet av $PM_{1(D)}$ modell	23
4.6	Fördelning av residualer korrigerade för autokorrelation och heteroskedasticitet av $PM_{1(D)}$ modell	23
4.7	Skattade värden korrigerade för autokorrelation och heteroskedasticitet av $PM_{1(D)}$ modell	25
4.8	Skattade värden korrigerade för autokorrelation och heteroskedasticitet av $PM_{1(D)}$	28
4.9	Residualer över tid av $PM_{1-2.5(D)}$ modell	29
4.10	ACF och PACF för fastställande av GARCH för $PM_{1-2.5(D)}$	29
4.11	ACF och PACF för fastställande av Cochrane-Orcutt för $PM_{1-2.5(D)}$	30
4.12	Residualer över tid korrigerade för autokorrelation och heteroskedasticitet av $PM_{1-2.5(D)}$	31
4.13	ACF och PACF korrigerade för autokorrelation och heteroskedasticitet av $PM_{1-2.5(D)}$	32
4.14	Fördelning av residualer korrigerade för autokorrelation och heteroskedasticitet av $PM_{1-2.5(D)}$	32
4.15	Skattade värden korrigerade för autokorrelation och heteroskedasticitet av $PM_{1-2.5(D)}$	33
4.16	Skattade värden korrigerade för autokorrelation och heteroskedasticitet av $PM_{1-2.5(D)}$	37
4.17	Residualer över tid för $PM_{2.5-10(D)}$	38
4.18	ACF och PACF för fastställande av GARCH för $PM_{2.5-10(D)}$	38
4.19	ACF och PACF för fastställande av Cochrane-Orcutt för $PM_{2.5-10(D)}$	39
4.20	Residualer över tid korrigerade för autokorrelation och heteroskedasticitet av $PM_{2.5-10(D)}$	40
4.21	ACF och PACF korrigerade för autokorrelation och heteroskedasticitet av $PM_{2.5-10(D)}$	41
4.22	Fördelning av residualer korrigerade för autokorrelation och heteroskedasticitet av $PM_{2.5-10(D)}$	41

4.23	Skattade värden korrigerade för autokorrelation och heteroskedasticitet av PM _{2.5-10(D)}	42
4.24	Skattade värden korrigerade för autokorrelation och heteroskedasticitet av PM _{2.5-10(D)}	46

Tabeller

2.1	Beskrivande statistik för skillnaden mellan korsningen och väglänken . . .	7
2.2	Förklaringsvariabler	8
2.3	Beskrivande statistik för förklarande variabler	10
4.1	Skattning av $\alpha_0, \dots, \alpha_6$ för $PM_{1(D)}$ med tidsvariabler	21
4.2	Skattning av ϕ_1, ϕ_2 och ϕ_3 för $PM_{1(D)}$ med tidsvariabler	22
4.3	Skattning av β_0, β_1 , och β_2 för $PM_{1(D)}$ med tidsvariabler	24
4.4	Bakåteliminering för $PM_{1(D)}$	26
4.5	Partiellt F-test av vindriktning för $PM_{1(D)}$	26
4.6	Skattning av β_0, \dots, β_8 för $PM_{1(D)}$ med förklaringsvariabler	27
4.7	Skattning av α_0 , och α_1 för $PM_{1-2.5}$ med tidsvariabler	30
4.8	Skattning av ϕ_1, \dots, ϕ_5 för $PM_{1-2.5(D)}$ med tidsvariabler	31
4.9	Skattning av β_0, \dots, β_3 för $PM_{1-2.5(D)}$ med tidsvariabler	33
4.10	Bakåteliminering för $PM_{1-2.5(D)}$	34
4.11	Partiellt F-test av vindriktning för $PM_{1-2.5(D)}$	35
4.12	Skattning av $\beta_0, \dots, \beta_{10}$ för $PM_{1-2.5}$ med förklaringsvariabler	36
4.13	Skattning av α_0 , och α_1 för $PM_{2.5-10(D)}$ med tidsvariabler	39
4.14	Skattning av ϕ_1, \dots, ϕ_5 för $PM_{2.5-10(D)}$ med tidsvariabler	40
4.15	Skattning av β_0, \dots, β_3 för $PM_{2.5-10(D)}$ med tidsvariabler	42
4.16	Bakåteliminering för $PM_{2.5-10(D)}$	43
4.17	Partiellt F-test av vindriktning för $PM_{2.5-10(D)}$	44
4.18	Skattning av β_0, \dots, β_8 för $PM_{2.5-10(D)}$ med förklaringsvariabler	44
6.1	Partikelhalternas påverkan av förklaringsvariabler	50
A.1	Skattning av $\alpha_0, \dots, \alpha_6$ för $PM_{1(D)}$ med förklaringsvariabler	i
A.2	Skattning av ϕ_1, ϕ_2 och ϕ_3 för $PM_{1(D)}$ med förklaringsvariabler	i
A.3	Skattning av α_0 och α_6 för $PM_{1-2.5}$ med förklaringsvariabler	ii
A.4	Skattning av ϕ_1, \dots, ϕ_5 för $PM_{1-2.5}$ med förklaringsvariabler	ii
A.5	Skattning av α_0 och α_3 för $PM_{2.5-10(D)}$ med förklaringsvariabler	iii
A.6	Skattning av ϕ_1, \dots, ϕ_5 för $PM_{2.5-10(D)}$ med förklaringsvariabler	iii

Ordlista

PM₁	Partiklar som är mindre än 1 mikrometer i diameter.
PM_{1–2.5}	Partiklar som är mellan 1 till 2.5 mikrometer i diameter.
PM_{2.5}	Partiklar som är mindre än 2.5 mikrometer i diameter.
PM_{2.5–10}	Partiklar som är mellan 2.5 till 10 mikrometer i diameter.
PM₁₀	Partiklar som är mindre än 10 mikrometer i diameter.
Vägdamm	Partiklar som är synliga samt PM _{2.5–10} som uppkommer av exempelvis sandning av vägar.
Emissioner	Utsläpp från fordon som omfattar partiklar och gaser.
Kväveoxider	Gas som uppkommer av emissioner och är mindre än 1 mikrometer i diameter.

1. Introduktion

I detta kapitel beskrivs bakgrund om arbetet, uppdragsgivare, uppsatsens syfte och frågeställningar.

1.1 Bakgrund

Sedan 1980-talet bevisas det att ett samband finns mellan luftens halt av inandningsbara partiklar och olika symptom i befolkningen (Gustafsson, 2002). Partikelmåttet med en diameter på 10 mikrometer (PM_{10}) är ett vanligt mått att studera på grund av dess dokumenterade effekter på människor och miljö. Partikelmåttet med en diameter på 2.5 mikrometer ($PM_{2.5}$) visar dock på att bidra till hälsofarliga problem som till exempel astma och kroniska obstruktiva lungsjukdomar. Det är även påvisat att partiklar mindre än 2.5 mikrometer kan leda till hjärt- och kärlsjukdomar (Forsberg och Segerstedt, 2004).

Det finns ett antal faktorer som har en stor påverkan på mängden partiklar. Variabler som vägbanans fuktighet, vindhastighet, årstid och temperatur har en påverkan (Johansson, Norman och Gustafsson 2008). Olika gaser så som kvävedioxider har också en påverkan på människors hälsa och uppkommer bland annat från emissioner av fordon. Kväveoxider hamnar inom kategorin för partiklar med en diameter på 1 mikrometer eller mindre (PM_1) (Janhäll, 2005). Regeringen har år 2010 utfärdat lagar där det framkommer trösklar för kvävedioxid. I § 10 luftkvalitetsförordning (2010:477) framkommer det att genomsnittshalten inte får förekomma över en viss gräns under en viss tidsperiod (Miljö och Energidepartementet, 2010).

År 2002 kom Europaparlamentet och rådet (EU) med direktiv om att minska luftföroreningarna till en nivå som minimerar skadliga effekter på människors hälsa och på miljön som helhet. Beslutet tas om att förbättra övervakningen och bedömningen av luftkvaliteten, nedfallet av föroreningar och tillhandahålla information till allmänheten (Europaparlamentet och råd, 2008).

För PM_{10} finns det, enligt direktivet av EU, olika gränser som inte får överskridas ett visst antal gånger per år (Gustafsson, 2002). Flera svenska städer överskrider dock gränsvärdena i direktivet av EU och sedan år 2011 pågår det ett intensivt arbete med att sänka halterna av luftpartiklar i bland annat Stockholm där problemen är omfattande. En stor del av de partiklar som finns på gator omfattar vägdamm och avgaser från fordon (Gustafsson m. fl., 2016).

I Sverige existerar antagandet om att en skillnad inte finns för partikelhalten mellan korsningar och väglänkar. Statens väg- och transportforskningsinstitut (VTI) menar att korsningar har blivit ett allt mer intressant ämne när det gäller miljörelaterade frågor, frågan har uppkommit om det faktiskt finns en skillnad mellan korsningar och väglänkar gällande partikelhalt. Till exempel uppkommer intresset av att undersöka ifall ett övergångsställe

bör ligga i en korsning eller inte.

1.2 Tidigare studier

Inom området för skadliga partiklar och vägdamm har forskning av bland annat VTI och Stockholms Luft- och Bulleranalys (SLB) gjorts där Johansson, Norman och Gustafsson (2008) undersöker hur partikelhalten har förändrats med tiden.

De kom fram till att partikelhalterna mellan år 2013 och 2015 sjunker och att mängden damm på vägytan har, på alla studerade gator utom Hornsgatan, en sjunkande trend under dessa år. Säsongsbundna variationer för mängden damm på vägytan förekommer också med stora mängder på vinter och tidig vår och låga mängder i oktober och maj.

Janhäll (2005) menar i sin studie om luftföroreningar att partiklar som omfattar PM_{10} är så små att de kan vara kvar i atmosfären under flera månader, vilket ger dem möjligheten att färdas längre distanser. Det framgår också att PM_{10} inte heller korrelerar med $PM_{2.5}$, då $PM_{2.5}$ korrelerar bättre med PM_{10} . Större partiklar som $PM_{2.5}$ och PM_{10} påverkas mer av lokala effekter.

Samet m. fl. (2000) skattar en modell för sjukhusvistelser eller dödlighet som förklaras av PM_{10} halter. Data till forskningen grundas på 90 städer i USA där halten av PM_{10} mäts. Modellen baserades på att undersöka på hur halter av PM_{10} längre bak i tiden påverkar den aktuella tidens sjukhusvistelser eller dödlighet. Samet m. fl. (2000) tar hänsyn till att korrelation förekommer i modellen. Samet m. fl. (2000) kombinerar därför alla städers halter av PM_{10} för att få bort korrelationen mellan observationerna.

1.3 Uppdragsgivare

VTI bedriver flera samhällsmotiverade tillämpade forskningsprojekt och utredningar som kan tillhandahållas av allmänheten. VTI har sedan 1920-talet haft i uppdrag av Sveriges regering att bedriva den forskning och de utredningar som omfattar områden kring infrastruktur, trafik och transporter. Bland dessa utredningar och projekt bedrivs forskning rörande miljö och en del av dessa projekt är om halter och emissioner på väg-nivå (VTI, 2017).

1.4 Syfte

Syftet med rapporten är att undersöka ifall det finns en skillnad i partikelhalter mellan en korsning och en väglänk då man tidigare, i Sverige, har antagit att det inte finns en skillnad mellan en korsning och en väglänk.

Hur meteorologiska, trafikrelaterade och tidsrelaterade variabler påverkar dessa skillnader är också av intresse.

1.4.1 Frågeställningar

- Finns det någon signifikant skillnad över tid i halter av olika partiklar mellan korsningen på Hornsgatan 104 och väglänken på Hornsgatan 110?
- Har meteorologiska, trafikrelaterade och tidsrelaterade variabler någon påverkan på dessa skillnader i partikelhalter?

1.5 Etiska och samhälleliga aspekter

Uppsatsen utförs i avseende att upptäcka skillnader och därmed ny kunskap om partikelhalter i olika förhållanden. Förståelsen om partikelhalter gynnar staten likaså allmänheten där man försöker att få ner halterna i det svenska samhället. Uppsatsen kommer att gynna förståelsen om skillnader och underlätta eventuell samhällsplanering.

Uppsatsen innehåller inte någon information om enskilda individer eller grupper. Uppsatsen innehåller endast information om hur situationen på korsningen och väglänken ser ut med avseende på partikelhalter, meteorologiska och trafikrelaterade variabler.

2. Data

Kapitlet tar upp vilka variabler som används i uppsatsen och beskrivande statistik om dem. Variabler som benämns som reponsvariabler är variabler som uppsatsen har som syfte att förklara. För att förklara reponsvariablerna använder sig uppsatsen av andra variabler som benämns som förklaringsvariabler.

Datamaterialet som erhålls samlas in via tre stycken mätstationer. Partiklarna PM_1 , $PM_{2.5}$ och PM_{10} mäts i korsningen och på väglänken medan meteorologiska variabler mäts på Torkel knutssongata. Alla variabler mäts med femtonminuters mellanrum från och med den första mars 2016 klockan 13:15 till och med den första maj 2016 klockan 00:00. Varje variabel har därför totalt 5802 stycken mätpunkter.

Partiklar som tillhör PM_1 omfattar väldigt fint damm, förbränningspartiklar och gaser. Partiklar som tillhör $PM_{2.5}$ omfattar PM_1 och lite större partiklar som sporer och andra organiska partiklar. PM_{10} omfattar $PM_{2.5}$ och ännu större partiklar som rök och grövre fint damm (Camfil, 2012). PM -värdena är därmed delmängder av varandra ($PM_1 \subset PM_{2.5} \subset PM_{10}$).

2.1 Responsvariabler

För att undersöka de enskilda partikelhalterna görs PM -variablerna så att varje halt är ett separat intervall. För att få separata intervall av partikelhaterna utförs beräkningen

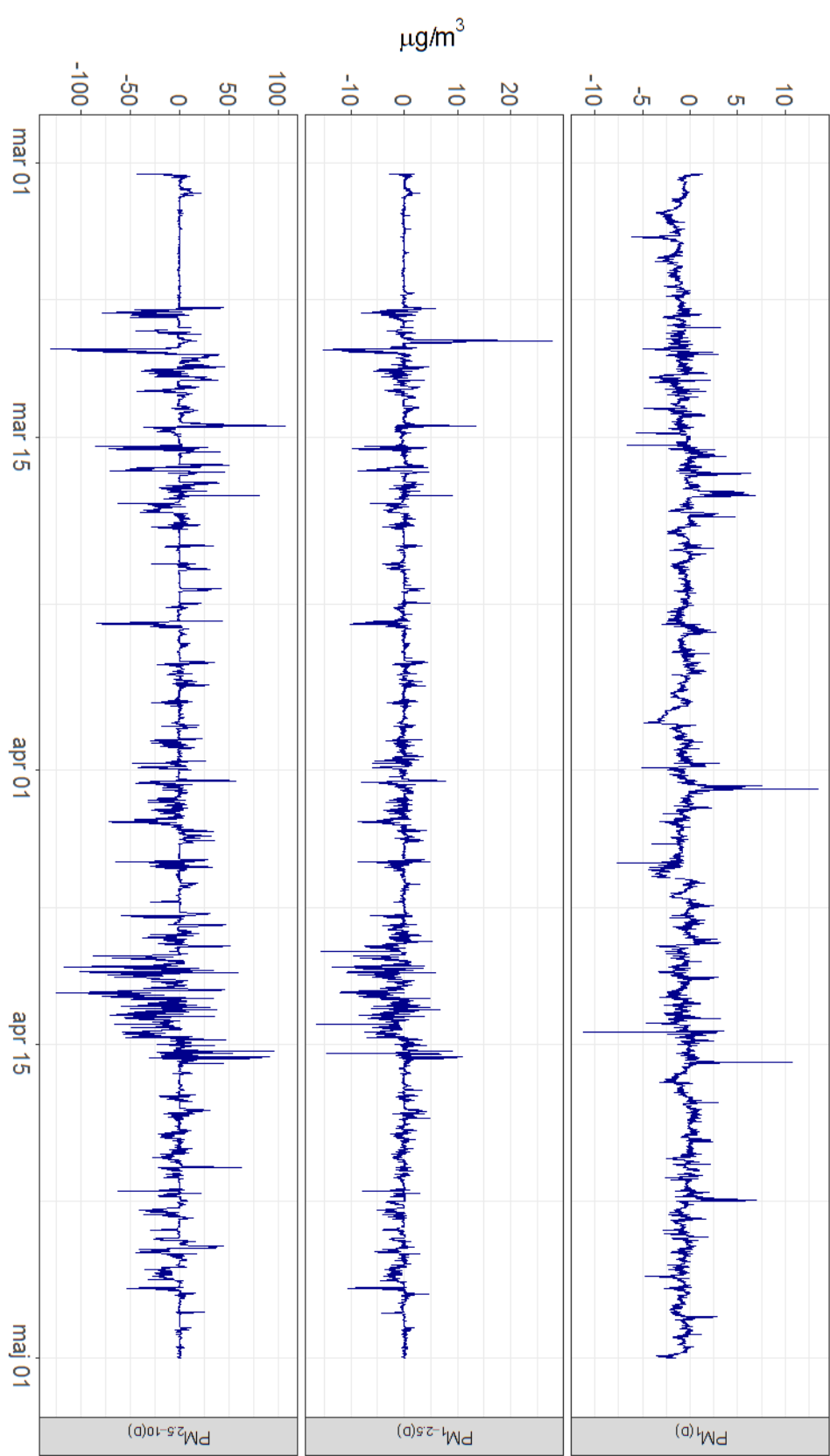
$$\begin{aligned} PM_{1-2.5} &= PM_{2.5} - PM_1 \\ PM_{2.5-10} &= PM_{10} - PM_{2.5} \end{aligned}$$

Differentiering av de separata PM -intervallen utförs mellan de två mätstationerna för att studera skillnaden mellan korsningen och väglänken. Differentieringen utförs

$$\begin{aligned} PM_{1(D)} &= PM_{1,Korsningen} - PM_{1,Väglänken} \\ PM_{1-2.5(D)} &= PM_{1-2.5,Korsningen} - PM_{1-2.5,Väglänken} \\ PM_{2.5-10(D)} &= PM_{2.5-10,Korsningen} - PM_{2.5-10,Väglänken} \end{aligned}$$

där $PM_{1(D)}$, $PM_{1-2.5(D)}$ och $PM_{2.5-10(D)}$ är responsvariablerna som analyseras i resultatkapitlet. Figur 2.1 visar hur responsvariablerna ser ut över tid.

Skilnaden mellan korsningen och väglänken



Figur 2.1: Responsvariablerna över tid

Beskrivande statistik av responsvariablerna presenteras i figur 2.1 där min och max visar det minsta samt det största värdet variabeln har uppnått under den observerade perioden. I kolumnen NA (not available) ses antalet saknade observationer som variabeln innehåller.

Tabell 2.1: Beskrivande statistik för skillnaden mellan korsningen och väglänken

Skillnaden mellan korsningen och väglänken					
Variabel	Medelvärde	Standardavvikelse	Min	Max	NA
$PM_{1(D)}$	-0.6631	1.171	-11.209	26.117	38
$PM_{1-2.5(D)}$	-0.3800	1.880	-16.372	27.670	38
$PM_{2.5-10(D)}$	-2.289	15.225	-130.55	107.69	38

Det kan konstateras att väglänken generellt har högre partikelhalter än korsningen då samtliga medelvärden är negativa i tabell 2.1. Min och max påvisar extremvärden i data då de ligger så långt ifrån medelvärdet givet den låga standardavvikelsen, vilket även ses i figur 2.1. Figuren visar på att heteroskedasticitet förekommer i data.

2.2 Förklaringsvariabler

Tabell 2.2 visar de förklaringsvariabler som tillhandahålls för att undersöka ifall de har ett samband med responsvariablerna.

Variabler som *temperatur*, *relativ fuktighet*, *lufttryck*, *maxvind*, *vindriktning* och *nederbörd* klassas som meteorologiska variabler. Som tidigare nämnts mäts de meteorologiska variablerna på en mätstation på Torkel Knutssonsgata. Gatan ligger ungefär 500 meter ifrån korsningen och väglänken. De tre mätstationerna är så pass nära varandra så att samma väderförhållanden kan antas råda. Torkel knutssonsgata kan därmed användas för att förklara väderförhållandena för korsningen och väglänken.

Variabeln *trafik* indikerar om det är hög eller låg trafik. Variabeln mäts genom att studera hur många bilar som passerar väglänken.

Variablerna *tid*, *månad* och *periodicitet* är till för att förklara ifall en tidstrend finns i responsvariablerna. Variabeln *tid* förklarar ifall det finns en allmän trend i data medan *månad* används för att studera ifall månad mars skiljer sig mot månad april. Variabeln *periodicitet* skapas utifrån att figur 2.2 visar ett tydligt mönster för responsvariablerna $PM_{1-2.5(D)}$ och $PM_{2.5-10(D)}$. Mellan klockan 9:00 och 18:00 tyder figuren på att större negativa värden förekommer i jämförelse med resterande tidpunkter. En sinuskurva skapas därmed för att fånga upp detta mönster (Box m. fl., 2016). Då mönstret inte förekommer för responsvariabeln $PM_{1(D)}$ används inte variabeln *periodicitet* på den responsvariabeln. I figur 2.2 har sinuskurvan olika amplitud beroende på vilken responsvariabel den förklarar,

vilket möjliggör en bättre visualisering av hur variabeln *periodicitet* skulle kunna hjälpa till att förklara responsvariablerna. I kapitel 4 används variabeln i originalform, vilket innebär att variabeln inte justeras för varje responsvariabel. Hur variabeln ser ut över tid ses i figur 2.3. För att skapa variabeln används formel

$$Periodicitet_t = \sin(x_t \cdot \frac{\pi \cdot 4.75}{180})$$

där

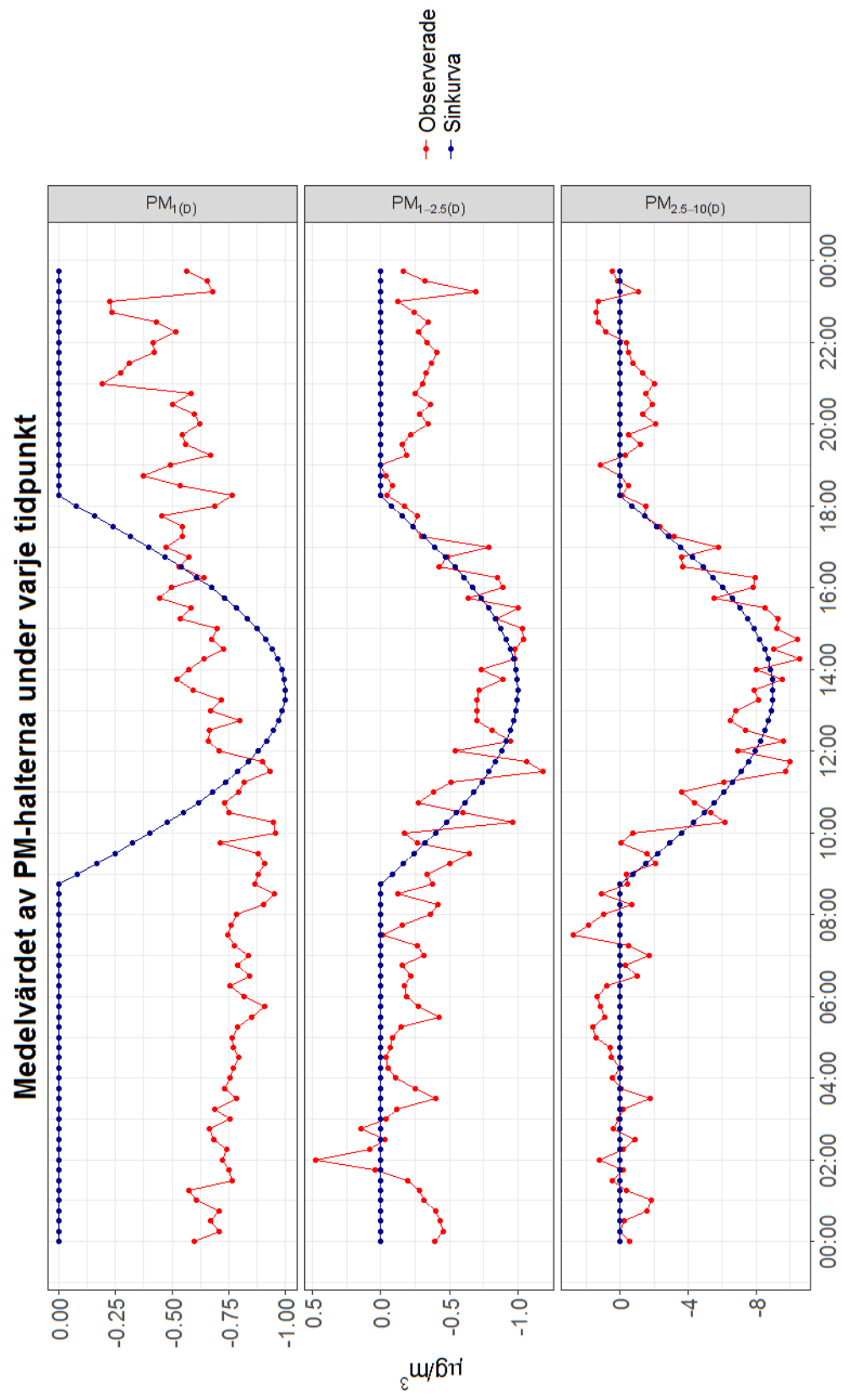
$$x_t = \begin{cases} 0, & \text{om klockan är innan 9:00 och efter 18:00} \\ 1, & \text{om klockan är 9:00} \\ \dots & \dots \\ 37, & \text{om klockan är 18:00} \end{cases}$$

Kolumnen skala i tabell 2.2 visar vilken skala variablerna har i modellerna i kapitel 4. Att variabeln är standardiserad innebär att variabeln fått ett medelvärde på noll och en standardavvikelse på ett.

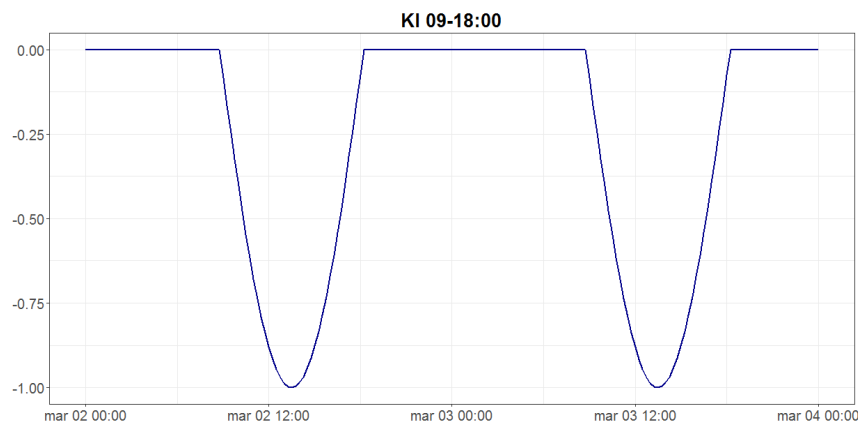
Tabell 2.2: Förklaringsvariabler

Förklarande variabler i uppsatsen

Variabel	Förklaring	Skala
Tid	En trend i data.	Värden 0-5800
Månad	Vid mars är det 0 och vid april är det 1.	Värden 0 och 1
Periodicitet	Förklarar periodiciteten som förekommer mellan klockan 09:00-18:00.	Värden mellan 0 och 1
Relativ fuktighet	Andelen vattenånga i förhållande till den maximala mängden vattenånga vid aktuell temperatur.	Standardiserad
Maxvind	Maxvinden mätt i meter per sekund (m/s).	Standardiserad
Trafik	Antalet fordon som har passerat väglänken. Indikerar på hög / låg trafik.	Standardiserad
Temperatur	Temperatur mätt i celsius.	Standardiserad
Lufttryck	Lufttrycket mätt i hektopascal (hPa).	Standardiserad
Nederbörd	Ifall det har varit nederbörd är det 1 annars 0.	Värden 0 och 1
Vindriktning longitud	Vid negativa värden är det västlig vind och vid positiva östlig vind.	Standardiserad
Vindriktning latitud	Vid negativa värden är det sydlig vind och vid positiva nordlig vind.	Standardiserad
Vindriktning interaktion	Interaktion för vindriktningen i latitud och longitud.	Standardiserad



Figur 2.2: Medelvärdet av klockslag för varje dag med sinuskurva



Figur 2.3: Sinuskurva för kl 09-18:00

Tabell 2.3: Beskrivande statistik för förklarande variabler

Förklarande variabler					
Variabel	Medelvärde	Standardavvikelse	Min	Max	NA
Relativ fuktighet	75.8	17.62	22.93	100.1	29
Maxvind	5.14	2.39	0.2	30.04	67
Temperatur	4.4	3.09	-2.55	13.28	29
Lufttryck	1005.5	9.71	983.8	1030.0	4
Trafik	194.22	108.6	13.0	466.0	4
Vindriktning longitud	-0.5	2.39	-19.66	10.66	40
Vindriktning latitud	-0.59	2.51	-10.47	13.49	40

Figur 2.3 visar beskrivande statistik för de kontinuerliga förklaringsvariablerna. Det konstateras att bara ett fåtal saknade värden finns i data jämfört med storleken på datamaterialet. Precis som för responsvariablerna i tabell 2.1, visar min och max det minsta och det största värdet som variabeln uppnår under den observerade perioden.

2.3 Imputering

Saknade värden förekommer i datamaterialet på grund av att underhåll av mätstationerna genomförs med jämna mellanrum samt att tekniska fel kan uppstå. Respons- och förklaringsvariablerna som används i uppsatsen innehåller få saknade värden i förhållande till antalet observerade värden, vilket visas i tabell 2.1 och 2.3. Metoden för imputering som

utförs är en typ av medelvärdesimputering. Om det till exempel finns ett saknat värde kl 12:00 en onsdag tas medelvärdet av alla onsdagar kl 12:00. Medelvärdet som erhålls imputerar man sedan med (Donders m.fl., 2006).

3. Metod

Kapitlet tar upp vilka metoder som används för att analysera data och besvara frågeställningarna.

3.1 Tidsserieregression

För att undersöka ifall förklarande variabler har en påverkan på responsvariabeln kan en multipel linjär regressionsmodell över tid anpassas (Wooldridge, 2011). En sådan modell följer strukturen i ekvation 3.1.

$$y_t = \beta_0 + \beta_1 x_{1t} + \dots + \beta_m x_{mt} + \epsilon_t \quad (3.1)$$

Modellen som anpassas i ekvation 3.1 har m stycken förklarande variabler där β_i , $i = 1, 2, \dots, m$ är parametern för varje förklaringsvariabel och där ϵ_t är feltermen för modellen. En sådan modell måste uppfylla sex stycken krav (Wooldridge, 2011).

1. Regressionen är linjär, vilket innebär att modellen kan skrivas med ekvation 3.1.
2. Förklaringsvariablerna får inte vara linjära kombinationer av varandra. Det får inte råda multikollinearitet.
3. Väntevärdet för feltermen måste vara noll för alla tidpunkter givet förklaringsvariablerna, $E(\epsilon_t | \mathbf{X}) = 0$.
4. Variansen för feltermen måste vara konstant för alla tidpunkter givet förklaringsvariablerna, $Var(\epsilon_t | \mathbf{X}) = \sigma^2$.
5. Det får inte råda autokorrelation i feltermen, $Corr(\epsilon_t, \epsilon_{t+k} | \mathbf{X}) = 0$.
6. Feltermen följer en normalfördelning med medelvärde noll och en konstant varians, $\epsilon_t \sim N(0, \sigma^2)$

I en multipel linjär regression över tid är det vanligt att krav nummer 4 och 5 inte är uppfyllda (Wooldridge, 2011). Modellen kan därmed behöva anpassas så att dessa krav blir uppfyllda. För att korrigera för heteroskedasticitet i feltermen har exempelvis Chen m. fl. (2015) finansiell data över tid där heteroskedasticitet är ett problem, vilket är krav nummer 4. Feltermen modelleras för att korrigera för heteroskedasticiteten.

3.1.1 Modellering för heteroskedasticitet

Där $Var(\epsilon_t | \mathbf{X}) = \sigma^2$ inte råder i feltermen för ekvation 3.1, kan problemet med heteroskedasticitet finnas och behöver därmed modelleras.

Modellering för heteroskedacitet i feltermen kan ske med en autoregressiv- eller glidande medelvärdesstruktur, eller med båda strukturerna i samma modellering. En modell som kan ta hänsyn till detta är en GARCH(s, r) på feltermen där s är autoregressivstrukturen och r är medelvärdesstrukturen. För att avgöra ordningen s och r i GARCH studeras autokorrelationsfunktionen (ACF) och partiella autokorrelationsfunktionen (PACF). ACF förklarar relationen mellan observationen t som är separerad med k tidsförskjutningar (lag). Partiella autokorrelationsfunktionen (PACF) förklarar den partiella korrelation mellan observationen t som är separerad med k lag (Box m. fl., 2016; Bowerman, O'Connell och Koehler, 2004).

Beräkning av ACF och PACF görs med ekvation 3.2 respektive 3.3.

$$r_k = \frac{\sum_t^{T-k} (\hat{\epsilon}_t^2 - \bar{\epsilon}^2)(\hat{\epsilon}_{t+k}^2 - \bar{\epsilon}^2)}{\sum_t^n (\hat{\epsilon}_t^2 - \bar{\epsilon}^2)}, \quad k = 0, 1, 2, \dots \quad (3.2)$$

där $\hat{\epsilon}_t$ är feltermen för observation t .

$$r_{kk} = \begin{cases} \frac{r_k - \sum_{j=1}^{k-1} r_{k-1,j} r_{k-j}}{1 - \sum_{j=1}^{k-1} r_{k-1,j} r_j}, & \text{om } k = 2, 3, \dots \\ r_1, & \text{om } k = 1 \end{cases} \quad (3.3)$$

där

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

Autokorrelationsfunktionen är r_k för lag k och r_{kk} är partiella autokorrelationsfunktionen för lag k .

Om ett avtagande mönster inträffar i PACF och r stycken spikar i ACF bör modellen korrigeras med en glidande medelvärdesstruktur med ordningen r . Däremot om ett avtagande mönster inträffar i ACF och s stycken spikar i PACF ska modellen modelleras med en autoregressiv struktur med ordning s (Box m. fl., 2016).

Då endast problematiken med att ett avtagande mönster i ACF förekommer och s stycken spikar i PACF, så presenteras endast GARCH($s, 0$).

Feltermen antas ifrån den multipla linjära regressionsmodellen som presenteras i ekvation 3.1 att följa

$$\epsilon_t = \sqrt{\sigma_t^2} u_t$$

där u_t är den homoskedastiska feltermen, och

$$\sigma_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \dots + \alpha_s \epsilon_{t-s}^2 \quad (3.4)$$

för s stycken lag och där $\alpha_0, \alpha_1, \dots, \alpha_s$ är parameterar som är med och förklarar heteroskedaciteten. För att skatta modellen i ekvation 3.4 används

$$\hat{\epsilon}_t^2 = \alpha_0 + \alpha_1 \hat{\epsilon}_{t-1}^2 + \dots + \alpha_s \hat{\epsilon}_{t-s}^2 + w_t$$

där w_t är feltermen för modellen. Kraven som inte får inträffa är $\sum_{i=1}^s \alpha_i < 1$ samt att $\alpha_0 > 0$ måste gälla för att modellen ska vara giltig (Box m. fl., 2016; Wooldridge, 2011).

Modellen som slutligen skattas med hjälp av GARCH($s,0$) följer strukturen

$$y_t / \sqrt{\sigma_t^2} = \beta_0 (1 / \sqrt{\sigma_t^2}) + \beta_1 (x_{1t} / \sqrt{\sigma_t^2}) + \dots + \beta_m (x_{mt} / \sqrt{\sigma_t^2}) + u_t \quad (3.5)$$

där feltermen i denna modell följer strukturen $u_t = \epsilon_t / \sqrt{\sigma_t^2}$.

Modellen som presenteras i ekvation 3.5 korregerar för krav nummer fyra i den multipla linjära regressionsmodellen som presenteras i ekvation 3.1 (Wooldridge, 2011).

3.1.2 Modellering för autokorrelation

Förekommer autokorrelation i feltermen för ekvation 3.5, $\text{Corr}(u_t, u_{t+k} | \mathbf{X}) \neq 0$, behöver feltermen (u_t) modelleras. Detta kan göras genom att lägga till en ARMA(p, q) struktur på feltermen där p är den autoregressiva strukturen och q är glidande medelvärdesstrukturen (Wooldridge, 2011; Cochrane och Orcutt, 1949).

För att avgöra hur många steg bakåt i tiden autokorrelation råder samt hur feltermen behöver modelleras studeras ACF och PACF på feltermen. Beräkning av ACF och PACF görs med ekvation 3.2 respektive 3.3 och istället för $\hat{\epsilon}^2$ så beräknas ekvationerna på \hat{u}_t . Förekommer ett avtagande mönster i PACF och q stycken spikar i ACF bör modellen modelleras med en glidande medelvärdesstruktur på feltermen. Förekommer däremot ett avtagande mönster i ACF och p stycken spikar i PACF bör modellen modelleras med en autoregressiv struktur på feltermen (Wooldridge, 2011; Box m. fl., 2016).

Då endast problematiken med att ett avtagande mönster i ACF förekommer och p stycken spikar i PACF, så presenteras endast en autoregressiv struktur med ordningen p . Skattningarna i den autoregressiva strukturen för feltermen modelleras med metoden Cochrane-Orcutt. Metoden baserar sig på en upprepande process för att få fram parameterskattningar som inte skattar autokorrelerade feltermer.

När ordningen p har konstaterats skattas $\phi_1, \dots, \phi_p | \beta_0, \beta_1, \dots, \beta_m$ där $\beta_0, \beta_1, \dots, \beta_m$ skattas med hjälp av ekvation 3.5. För att kunna starta en upprepande process måste man ha en startpunkt och det får man genom att skatta ϕ_1, \dots, ϕ_p en gång på feltermen. Dessa ϕ_1, \dots, ϕ_p kommer man sedan att kunna använda för att starta den upprepande processen och få fram de slutgiltiga parameterskattningarna som tar hänsyn till autokorrelation. Ekvationen för att få fram ϕ_1, \dots, ϕ_p är

$$u_t = \phi_1 u_{t-1} + \phi_2 u_{t-2} + \dots + \phi_p u_{t-p} + a_t \quad (3.6)$$

där $\sum_{i=1}^p \phi_i < 1$ inte får inträffa och a_t är en felterm och där u_t är feltermen ifrån ekvation 3.5 (Box m. fl., 2016).

Från ekvation 3.6 används parameterskattningarna, ϕ_1, \dots, ϕ_p i en upprepande process för att transformera data och att beräkna följande steg:

1. Skatta $\beta_0, \beta_1, \dots, \beta_m | \phi_1, \phi_2, \dots, \phi_p$ där modellen blir

$$y_t^* = \beta_0 x_{0t}^* + \beta_1 x_{1t}^* + \dots + \beta_m x_{mt}^* + a_t \quad (3.7)$$

där

$$x_{0t}^* = 1 - \phi_1 - \dots - \phi_p$$

$$y_t^* = y_t - \phi_1 y_{t-1} - \dots - \phi_p y_{t-p}$$

$$x_{it}^* = x_{it} - \phi_1 x_{i,t-1} - \dots - \phi_p x_{i,t-p}$$

2. Ta fram nya feltermen $u_t | \beta_0, \beta_1, \dots, \beta_m$ där $\beta_0, \beta_1, \dots, \beta_m$ skattas av ekvation 3.7 och sedan skatta nya $\phi_1, \phi_2, \dots, \phi_p$ på den nya feltermen $u_t | \beta_0, \beta_1, \dots, \beta_m$. Skattningarna $\phi_1, \phi_2, \dots, \phi_p$ skattas med samma struktur som ekvation 3.6.

Steg 1 och 2 upprepas tills parametrarna ϕ_1, \dots, ϕ_p och β_0, \dots, β_m bara förändras i den sjunde decimalen (Wooldridge, 2011; Cochrane och Orcutt, 1949).

3.1.3 Bakåteleminering

Bakåteleminering utförs på en modell där alla potentiella förklarande variabler ingår där man undersöker ifall det finns variabler som inte är signifikanta i modellen. En signifikansnivå (α) väljs innan bakåteleminering börjar för att undersöka vilka variabler som inte bör vara med i modellen.

Metoden går ut på att ta bort en variabel åt gången. Variabeln som har det största p -värdet samt överstiger signifikansnivån tas bort från modellen och sedan skattas en ny modell på de resterande variablerna. Detta utförs tills alla förklarande variabler är signifikanta (Bowerman, O'Connell och Koehler, 2004).

3.1.4 Partiellt F-test

För att testa om en grupp variabler, givet en viss signifikansnivå (α), tillför något till en modell kan partiellt F-test användas. Testet grundar sig i att testa ifall den $g+1$:te variabeln till den m :te variabeln tillför något till modellen.

$$y_t = \beta_0 + \beta_1 x_{1t} + \dots + \beta_g x_{gt} + \dots + \beta_m x_{mt} + a_t$$

Hypoteserna är följande:

$$H_0 : \beta_{g+1} = \beta_{g+2} = \dots = \beta_m = 0$$

$$H_a : \text{Minst en av } \beta_{g+1}, \beta_{g+2}, \dots, \beta_m \text{ är inte lika med } 0$$

Den partiella F-statistikan är

$$F = \frac{(SSE_R - SSE_C)/(m - (g + 1))}{SSE_C/[n - (m + 1)]}$$

där n är antalet observationer i data. För den reducerade modellens felterm är SSE_R variansen och SSE_C variansen för den fulla modellens felterm.

Statistikan F jämförs mot det kritiska värdet $F_\alpha[(m - (g + 1)); (n - (m - (g + 1)))]$. Ifall $F > F_\alpha$ kan H_0 förkastas och man kan konstatera att minst en av $\beta_{g+1}, \beta_{g+2}, \dots, \beta_m$ är inte lika med 0, man kan således behålla gruppen av variabler i modellen. (Bowerman, O'Connell och Koehler, 2004).

3.2 Implementering i R

För att ha flexibilitet med metoderna används programspråket R i programvaran RStudio version 3.3.2 för alla statistiska beräkningar och visualiseringar.

3.2.1 Paket

Paket som används i uppsatsen är:

- ggplot2: för visualisering av data i uppsatsen.
- TSA: för att få bort spiken på ACF på lag 0.
- reshape: för att kunna omformatera data.

Metoderna för GARCH och Cochrane-Orcutt har kodats för hand och presenteras i bilaga B. För Cochrane-Orcutt har en funktions skapats där antalet ϕ kan bestämmas själv.

4. Resultat

Resultaten som presenteras i detta kapitel är baserade på metoder i kapitel 3. Tre separata analyser har genomförts, en för vardera responsvariabel där en signifikansnivå på fem procent valts ($\alpha = 0.05$). Dessa responsvariabler presenteras i figur 2.1 och tabell 2.1.

4.1 Analys av $PM_{1(D)}$

Responsvariabeln $PM_{1(D)}$ analyseras genom två separata multipla linjära tidsseriereggressionsmodeller. Den första modellen undersöker ifall det förekommer någon signifikant skillnad över tid av PM_1 mellan korsningen och väglänken. Den andra modellen undersöker ifall meteorologiska, trafikrelaterade och tidsrelaterade variabler har någon påverkan på skillnaden av PM_1 mellan korsningen och väglänken.

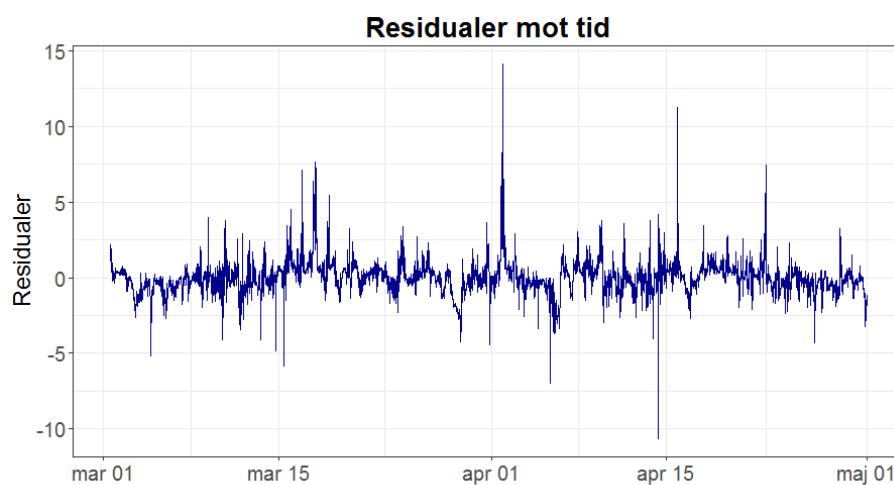
4.1.1 Modell med tidsvariabler

För att analysera ifall det förekommer någon signifikant skillnad av PM_1 över tid mellan korsningen och väglänken analyseras en multipel linjär tidsseriereggression med strukturen ifrån ekvation 3.1. Endast tidsrelaterade variabler förekommer i modellen så som *tid* och *månad* som presenteras i tabell 2.2. Anledningen till att variabeln *periodicitet* inte tas med i modellen är för att figur 2.2 påvisar att variabeln inte följer responsvariabeln $PM_{1(D)}$.

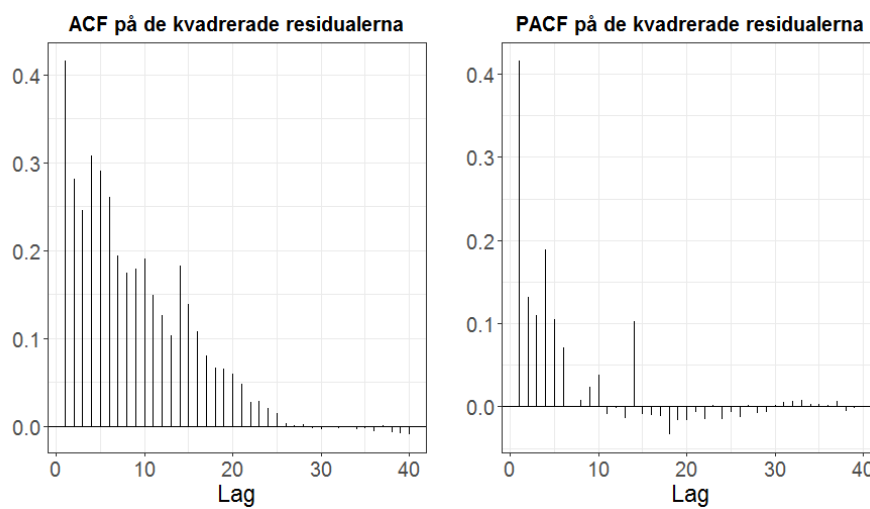
För att undersöka ifall modellens krav, som presenteras i kapitel 3.1, är uppfyllda studeras modellens residualer.

Figur 4.1 tyder på att residualerna ifrån modellen med strukturen ifrån ekvation 3.1 inte har en konstant varians över tid. För att modellera för en icke konstant varians byggs en GARCH(s, r) vidare på modellen. För att avgöra ordningen s och r studeras ACF och PACF på de kvadrerade residualerna.

Figur 4.2 visar ett avtagande möster i ACF och sex stycken spikar konstateras vara tydliga i PACF. Ordning s utses till 6 och ordningen r utses till 0. En modell med strukturen i ekvation 3.5 skattas därmed.



Figur 4.1: Residualer över tid av $PM_{1(D)}$ modell

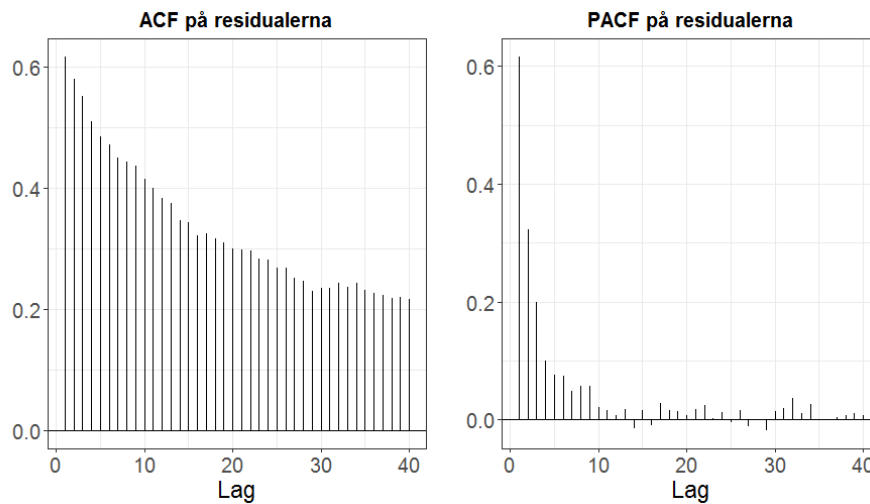


Figur 4.2: ACF och PACF för fastställande av GARCH för $PM_{1(D)}$

Tabell 4.1: Skattning av $\alpha_0, \dots, \alpha_6$ för $PM_{1(D)}$ med tidsvariabler

Modellering för heteroskedasticitet				
Parameter	Skattning	Standardavvikelse	t-värde	p-värde
α_0	0.369	0.061	6.071	< 0.001
α_1	0.287	0.013	21.882	< 0.001
α_2	0.065	0.014	4.760	< 0.001
α_3	0.040	0.013	2.965	0.003
α_4	0.146	0.013	10.831	< 0.001
α_5	0.087	0.014	6.374	< 0.001
α_6	0.073	0.013	5.564	< 0.001

Tabell 4.1 visar de parametrar som hjälper till att modellera för heteroskedasticitet. Parametrarna skattas med hjälp av ekvation 3.4. För att avgöra ifall modellen, som är korrigerad för heteroskedasticitet, uppfyller kraven studeras modellens residualer.



Figur 4.3: ACF och PACF för fastställande av Cochrane-Orcutt för $PM_{1(D)}$

Figur 4.3 visar att residualerna ifrån modellen med strukturen ifrån ekvation 3.5 innehåller en positiv autokorrelation, vilket ska korrigeras för. Ett avtagande mönster ses i ACF och tydliga spikar ses i PACF. Modellen ska därför modelleras med en autoregressiv struktur, vilket görs med Cochrane-Orcutt som presenteras i ekvation 3.7. Tre stycken spikar konstateras vara tydliga i PACF och p bestäms till 3.

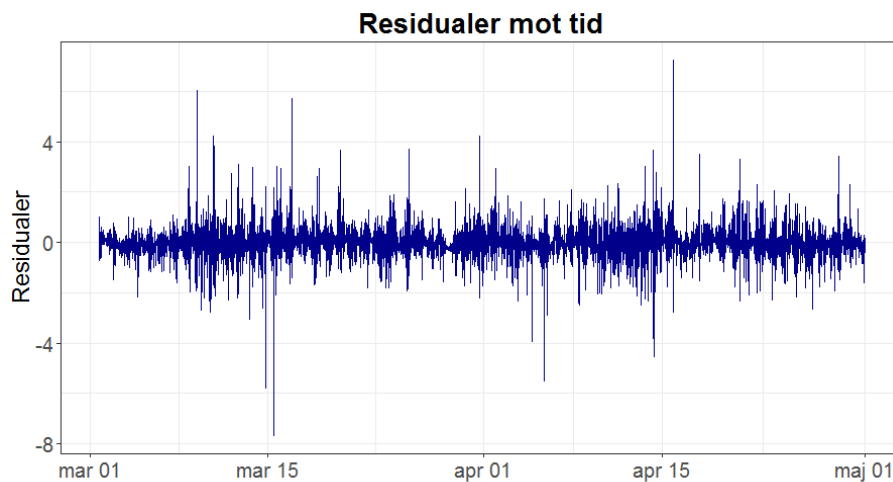
Tabell 4.2: Skattning av ϕ_1, ϕ_2 och ϕ_3 för $PM_{1(D)}$ med tidsvariabler

Modellering för autokorrelation				
Parameter	Skattning	Standardavvikelse	t-värde	p-värde
ϕ_1	0.353	0.013	27.436	< 0.001
ϕ_2	0.239	0.013	17.970	< 0.001
ϕ_3	0.201	0.013	15.590	< 0.001

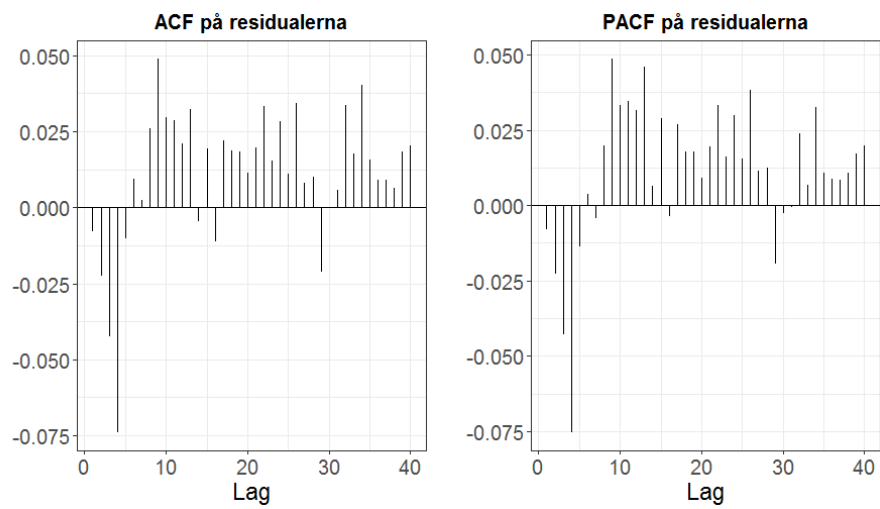
Tabell 4.2 visar parametrarna som hjälper till att modellera för autokorrelation. Parametrarna skattas med hjälp av ekvation 3.6.

Efter sju iterationer förändras inte parameterskattningarna av ϕ_1, ϕ_2 och ϕ_3 nämnvärt. Framtagning av parametrar för de tidsrelaterade förklaringsvariablerna, β_0, β_1 och β_2 , sker därmed med sju iterationer.

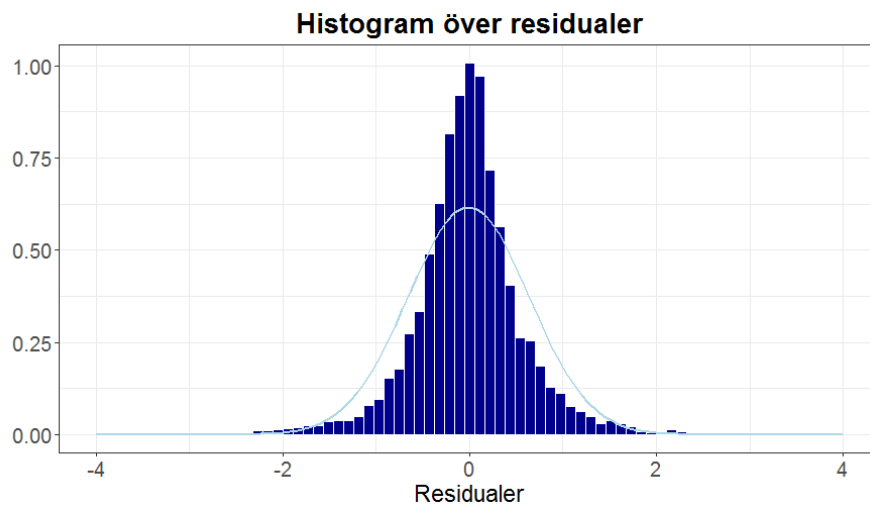
Innan β_0, β_1 och β_2 studeras, studeras modellens residualer för att se ifall modellen uppfyller dess krav.



Figur 4.4: Residualer över tid korrigerade för autokorrelation och heteroskedasticitet av $PM_{1(D)}$ modell



Figur 4.5: ACF och PACF korrigerade för autokorrelation och heteroskedasticitet av $PM_{1(D)}$ modell



Figur 4.6: Fördelning av residualer korrigerade för autokorrelation och heteroskedasticitet av $PM_{1(D)}$ modell

Figurerna 4.4 - 4.6 är baserade på residualerna ifrån modellen som tagit hänsyn till heteroskedasticitet och autokorrelation. Figur 4.4 tyder på att residualerna har en jämnare varians och figur 4.5 visar på att residualerna är okorrelerade. Figur 4.6 visar hur residualerna följer en normalfördelning, men att fördelningen är för toppig för att följa en normalfördelning. Dock för att gå vidare med modellen antas residualerna vara approximativt normalfördelade.

Modellen i tabell 4.3 har strukturen ifrån ekvation 3.7 och antas uppfylla kraven som presenteras i kapitel 3.1. Denna modell är modellerad för heteroskedasticitet genom en GARCH(6,0) samt modellering för autokorrelation sker med Cochrane Orcutt där $p=3$.

Tabell 4.3: Skattning av β_0, β_1 , och β_2 för $PM_{1(D)}$ med tidsvariabler

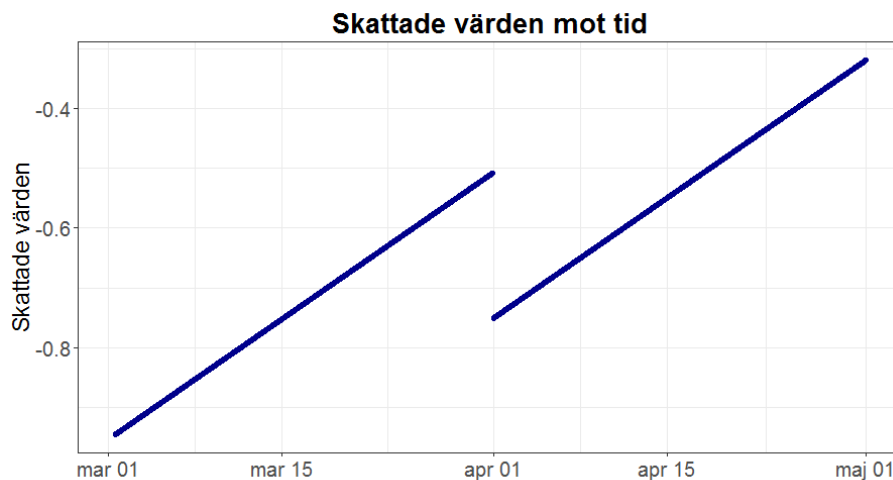
Parameterskattningar för $PM_{1(D)}$				
Variabel	Skattning (β_i)	Standardavvikelse	t-värde	p-värde
Intercept	-0.955	0.067	-14.326	< 0.001
Tid	0.00015	0.00001	4.249	< 0.001
Månad	-0.228	0.119	-1.998	0.046

Tabell 4.3 presenterar parameterskattningarna för interceptet och de tidsrelaterade förklaringsvariablerna som är skattade med modellen med strukturen ifrån ekvation 3.7.

Interceptet är signifikant skilt ifrån noll med en punktskattning på -0.955. Detta tolkas som att det finns en signifikant skillnad mellan korsningen och väglänken i början av tidsserien där den första skillnaden är skattad till -0.955. Vilket innebär att det förekommer generellt sett 0.955 mindre PM_1 i korsningen än på väglänken i början av tidsserien. Parametrarna för *tid* och *månad* är med i modellen, så skillnaden mellan korsningen och väglänken förväntas att ändras över tid.

Figur 4.7 visar de skattade värdena på responsvariabeln $PM_{1(D)}$ över tid som är framtagna med parameterskattningarna ifrån tabell 4.3. Figuren visar att den minsta generella skillnaden mellan korsningen och väglänken är skattad i slutet av tidsserien med en punktskattning på -0.319.

Under hela tidsserien är de skattade värdena under noll. Vilket tolkas som att det förekommer en generell skillnad mellan korsningen och väglänken under hela den observerade tidsserien där korsningen har mindre halt av PM_1 än väglänken.



Figur 4.7: Skattade värden korrigerade för autokorrelation och heteroskedasticitet av $PM_{1(D)}$ modell

4.1.2 Modell med förklaringsvariabler

För att analysera ifall meteorologiska, trafikrelaterade och tidsrelaterade variabler har någon påverkan på skillnaden av PM_1 mellan korsningen och väglänken skattas en multipel linjär tidsserieregression. Modellen korrigeras för heteroskedasticitet och autokorrelation genom GARCH($s = 6$, $r = 0$) och Cochrane-Orcutt där $p = 3$. Ordningen för s , r och p är samma som den modell som presenteras i kapitel 4.1.1. Detta är för att samma mönster i residualerna förekommer.

Förklaringsvariabler som får möjlighet att vara med i modellen är de som presenteras i tabell 2.2. Den enda förklaringsvariabel som inte är med från denna tabell i modellen är variabeln *periodicitet* då figur 2.2 påvisar att variabeln inte följer responsvariabeln $PM_{1(D)}$.

För att avgöra vilka förklaringsvariabler som signifikant påverkar responsvariabeln tillämpas bakåteliminering och ett partiellt F-test.

Tabell 4.4 visar att variablerna *temperatur*, *lufttryck* och *nerbörd* inte bör vara med och förklara responsvariabeln $PM_{1(D)}$. Modellen reduceras därmed genom att ta bort dessa förklaringsvariabler från modellen.

Tabell 4.4: Bakåteliminering för $PM_{1(D)}$

P-värden				
Variabel	Modell 1	Modell 2	Modell 3	Modell 4
Tid	< 0.001	< 0.001	< 0.001	< 0.001
Månad	0.001	0.001	0.001	0.001
Relativ fuktighet	< 0.001	< 0.001	< 0.001	< 0.001
Maxvind	< 0.001	< 0.001	< 0.001	< 0.001
Temperatur	0.216	0.219	0.240	-
Lufttryck	0.454	0.452	-	-
Trafik	0.072	0.068	0.084	0.042
Nerbörd	0.660	-	-	-

För att avgöra ifall vindriktning ska vara med i modellen genomförs ett partiellt F-test då vindriktning förklaras av tre stycken variabler.

Tabell 4.5: Partiellt F-test av vindriktning för $PM_{1(D)}$

Partiellt F-test					
Variabel	SSE	df	df residualer	F-värde	p-värde
Modell utan vindriktning	6638.105	5	5796		
Modell med vindriktning	5625.406	8	5793		
	1012.699	3		347.023	< 0.001

F-testet i tabell 4.5 visar att de tre variablerna som förklarar vindriktning bör vara med i modellen då p-värdet understiger signifikansnivån på fem procent.

När alla förklaringsvariabler som bör vara med i modellen enligt bakåteliminering och det partiella F-testet konstaterats skattas modellens parametrar.

Parametrarna som modellerar för heteroskedasticitet och autokorrelation kan ses i bilaga A.1

Efter sju iterationer förändras inte parameterskattningarna av ϕ_1, ϕ_2 och ϕ_3 nämnvärt. Framtagning av parametrar för förklaringsvariablerna, β_0, \dots, β_8 , sker därmed med sju iterationer.

Tabell 4.6: Skattning av β_0, \dots, β_8 för $PM_{1(D)}$ med förklaringsvariabler

Parameterskattningar för $PM_{1(D)}$				
Variabel	Skattning (β_i)	Standardavvikelse	t-värde	p-värde
Intercept	-0.892	0.060	-14.830	< 0.001
Tid	0.00001	0.00001	4.394	< 0.001
Månad	-0.300	0.106	-2.814	0.005
Relativ fuktighet	-0.142	0.026	-5.451	< 0.001
Maxvind	0.087	0.015	5.729	< 0.001
Trafik	-0.042	0.021	-2.022	0.043
Vindriktning longitud	-0.324	0.018	-17.712	< 0.001
Vindriktning latitud	-0.011	0.019	-0.575	0.565
Vindriktning interaktion	-0.105	0.012	-8.512	< 0.001

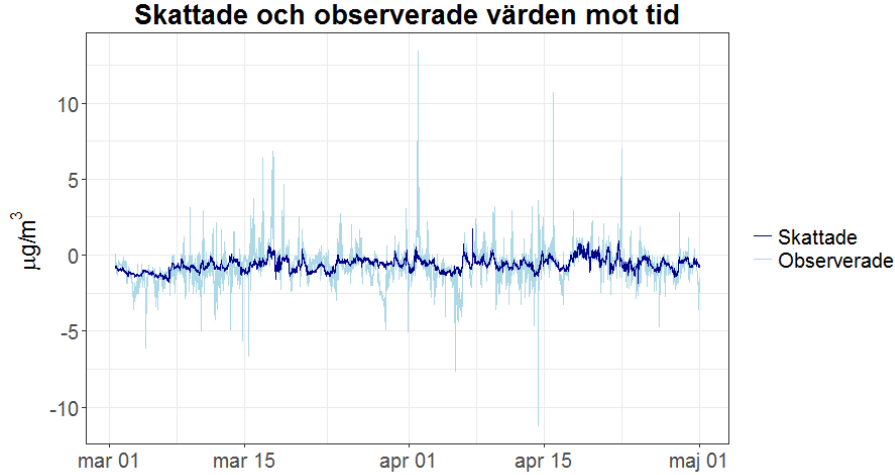
Tabell 4.6 presenterar parameterskattningarna för förklaringsvariablerna som är skattade med modellen med strukturen ifrån ekvation 3.7.

Tolkning av förklaringsvariabler givet att de andra förklaringsvariablerna hålls fixa:

- **Intercept, Tid, Månad dummy:** Den generella skillnaden av PM_1 mellan korsningen och väglänken förändras över tid givet att de andra förklaringsvariabler hålls på dess medelvärde.
- **Relativ fuktighet:** Är den relativa fuktigheten hög tenderar det till att mindre PM_1 förekommer i korsningen än på väglänken.
- **Maxvind:** Är maxvinden hög tenderar det till att mer PM_1 förekommer i korsningen än på väglänken.
- **Trafik:** Är det mycket trafik tenderar det till att mindre PM_1 förekommer i korsningen än på väglänken.
- **Riktning på vinden:** Vindriktning longitud har blivit skattad negativ vilket kan tolkas som att ifall det är östlig vind tenderar det till att mindre PM_1 förekommer i korsningen än på väglänken än om det hade varit västlig vind.
Vindriktning latitud är inte signifikant och bör därmed inte tolkas.
Vindriktning interaktion är signifikant vilket tyder på att det finns en interaktion mellan vindriktning longitud och vindriktning latitud.

Parametrar som är standardiserade är *relativ fuktighet*, *maxvind*, *trafik*, *vindriktning longitud*, *vindriktning latitud* och *vindriktning interaktion*. Då dessa variabler är standardiserade kan deras effekt på responsvariabeln $PM_{1(D)}$ jämföras. Variabeln *vindriktning*

longitud tyder på att ha störst påverkan på skillnaden mellan korsningen och väglänken då dess parameterskattning har störst absolutvärde.



Figur 4.8: Skattade värden korrigerade för autokorrelation och heteroskedasticitet av $PM_{1(D)}$

I figur 4.8 ses de skattade och observerade värdena för responsvariabeln $PM_{1(D)}$. Den skattade modellen fångar upp den generella trenden i responsvariabeln då de skattade värdena följer de observerade värdena relativt bra. Modellen har dock svårt att fånga upp extremvärden.

Förklaringsgraden, R^2 , för modellen är 22.35 procent.

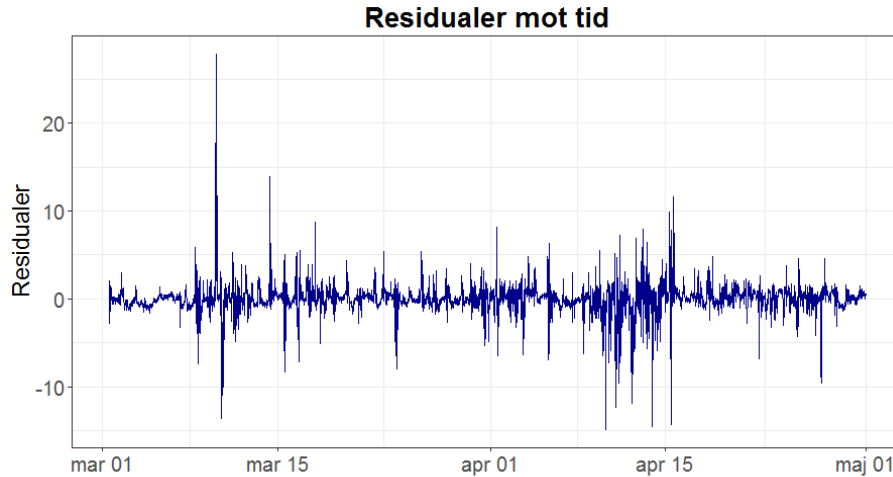
4.2 Analys av $PM_{1-2.5(D)}$

Responsvariabeln $PM_{1-2.5(D)}$ analyseras genom två separata multipla linjära tidsserie-regressionsmodeller. Den första modellen undersöker ifall det förekommer någon signifikant skillnad av $PM_{1-2.5}$ över tid mellan korsningen och väglänken. Den andra modellen undersöker ifall meteorologiska, trafikrelaterade och tidsrelaterade variabler har någon påverkan på skillnaden av halten $PM_{1-2.5}$ mellan korsningen och väglänken.

4.2.1 Modell med tidsvariabler

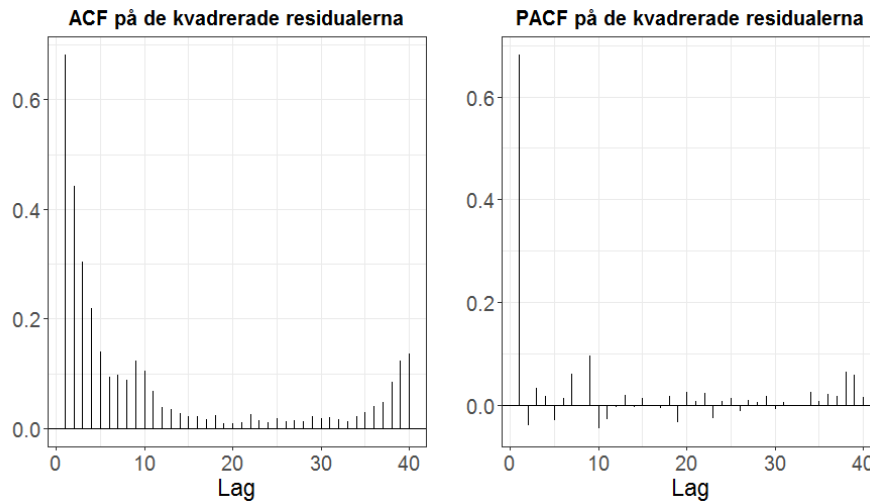
För att analysera ifall det förekommer någon signifikant skillnad av $PM_{1-2.5}$ över tid mellan korsningen och väglänken analyseras en multipel linjär tidsserieregression med strukturen ifrån ekvation 3.1. Endast tidsrelaterade variabler förekommer i modellen. Tidsrelaterade variabler som tas med i modellen är *tid*, *månad* och *periodicitet* som presenteras i tabell 2.2.

För att undersöka ifall modellens krav, som presenteras i kapitel 3.1, är uppfylla studeras modellens residualer.



Figur 4.9: Residualer över tid av $PM_{1-2.5(D)}$ modell

Figur 4.9 visar på att residualerna av modellen med strukturen ifrån ekvation 3.1 inte har en konstant varians över tid. För att modellera för en icke konstant varians byggs en $GARCH(s,r)$ vidare på modellen. För att avgöra ordningen s och r studeras ACF och PACF på de kvadrerade residualerna.



Figur 4.10: ACF och PACF för fastställande av GARCH för $PM_{1-2.5(D)}$

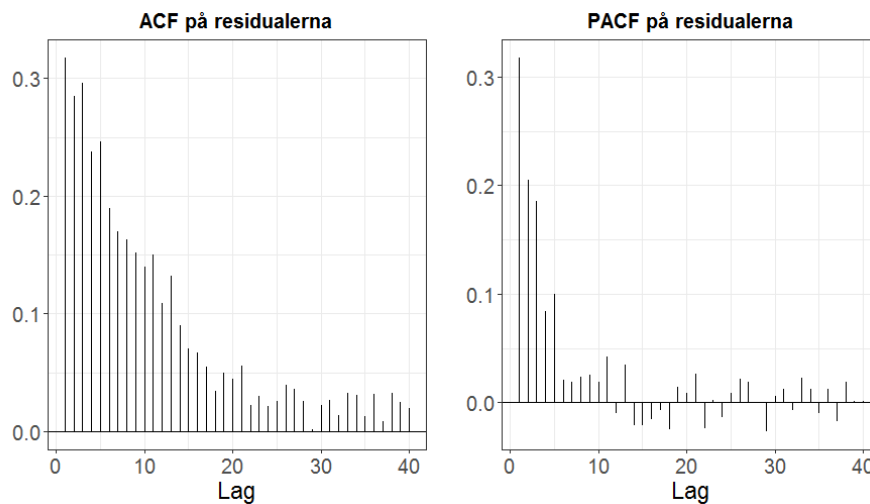
Figur 4.10 visar ett avtagande möster i ACF och en spik konstateras vara tydlig i PACF. Ordning s utses till 1 och ordningen r utses till 0. En modell med strukturen i ekvation 3.5 kan därmed skattas.

Tabell 4.7: Skattning av α_0 , och α_1 för $PM_{1-2.5}$ med tidsvariabler

Modellering för heteroskedasticitet				
Parameter	Skattning	Standardavvikelse	t-värde	p-värde
α_0	1.102	0.187	5.903	< 0.001
α_1	0.676	0.010	69.830	< 0.001

Tabell 4.7 presenteras de parametrar som hjälper till att modellera för heteroskedasticitet. Parametrarna skattas med hjälp av ekvation 3.4.

För att avgöra ifall modellen, som är korrigerad för heteroskedasticitet, uppfyller kraven studeras modellens residualer i figur 4.11.



Figur 4.11: ACF och PACF för fastställande av Cochrane-Orcutt för $PM_{1-2.5(D)}$

Figur 4.11 visar att residualerna ifrån modellen med strukturen ifrån ekvation 3.5 innehåller en positiv autokorrelation, vilket ska korrigeras för. Ett avtagande mönster visas i ACF och fem tydliga spikar visas i PACF. Modellen ska därför modelleras med en autoregressiv struktur, vilket görs med Cochrane-Orcutt som presenteras i ekvation 3.7. Fem stycken spikar konstateras vara tydliga i PACF och ordningen p bestäms till 5.

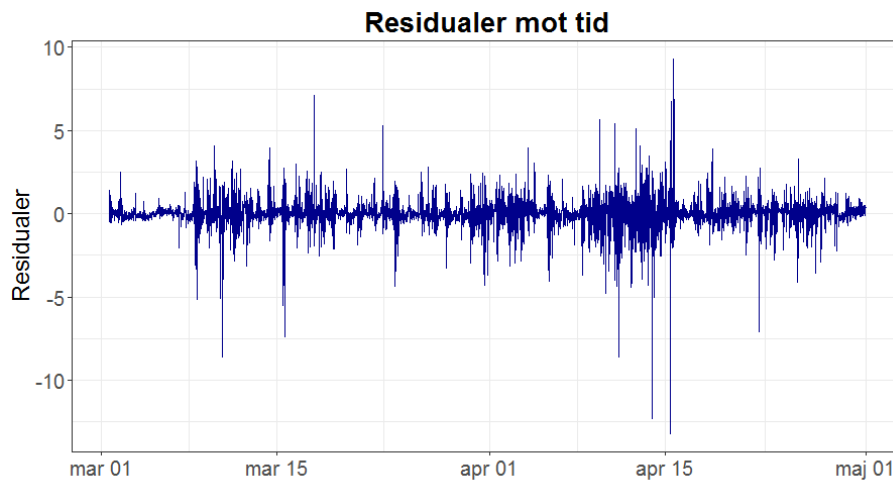
Tabell 4.8: Skattning av ϕ_1, \dots, ϕ_5 för $PM_{1-2.5(D)}$ med tidsvariabler

Modellering för autokorrelation				
Parameter	Skattning	Standardavvikelse	t-värde	p-värde
ϕ_1	0.188	0.013	14.428	< 0.001
ϕ_2	0.155	0.013	11.744	< 0.001
ϕ_3	0.160	0.013	12.155	< 0.001
ϕ_4	0.083	0.013	6.289	< 0.001
ϕ_5	0.131	0.013	10.044	< 0.001

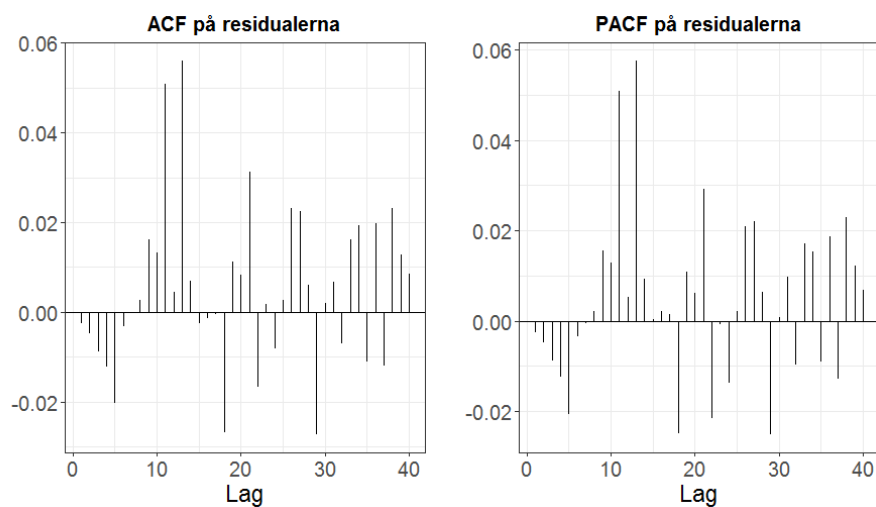
Tabell 4.8 visar parametrarna som hjälper till att modellera för autokorrelation. Parametrarna skattas med hjälp av ekvation 3.6.

Efter åtta iterationer förändras inte parameterskattningarna av ϕ_1, \dots, ϕ_4 nämnvärt. Framtagning av parametrar för de tidsrelaterade förklaringsvariablerna, β_0, \dots, β_3 , sker därmed med åtta iterationer.

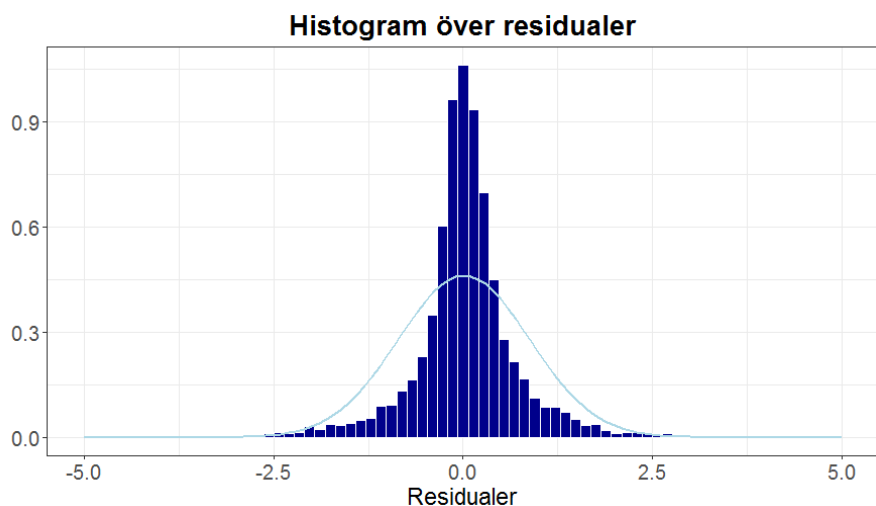
Innan β_0, \dots, β_3 studeras, studeras modellens residualer för att se ifall modellen uppfyller dess krav.



Figur 4.12: Residualer över tid korrigerade för autokorrelation och heteroskedasticitet av $PM_{1-2.5(D)}$



Figur 4.13: ACF och PACF korrigerade för autokorrelation och heteroskedasticitet av $PM_{1-2.5(D)}$



Figur 4.14: Fördelning av residualer korrigerade för autokorrelation och heteroskedasticitet av $PM_{1-2.5(D)}$

Figurerna 4.12 - 4.14 är baserade på residualerna ifrån modellen med strukturen i ekvation 3.7.

Figur 4.12 tyder på att residualerna har en jämnare varians och figur 4.13 visar på att residualerna är okorrelerade. Figur 4.14 visar hur residualerna följer en normalfördelning. Figuren visar att fördelningen är för toppig för att följa en normalfördelning. Dock för att gå vidare med modellen antas residualerna vara approximativt normalfördelade.

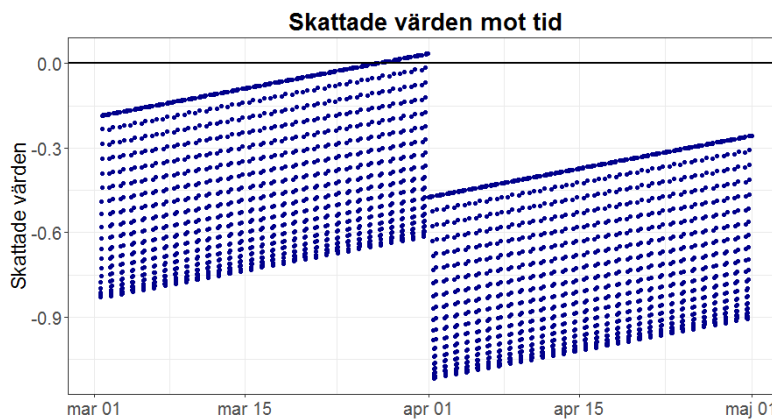
Modellen med strukturen ifrån ekvation 3.7 antas uppfylla dess krav. Denna modell är modellerad för heteroskedasticitet genom en GARCH(1,0) samt modellering för autokorrelation sker med Cochrane Orcutt där $p = 5$.

Tabell 4.9: Skattning av β_0, \dots, β_3 för $PM_{1-2.5(D)}$ med tidsvariabler

Parameterskattningar för $PM_{1-2.5(D)}$				
Variabel	Skattning	Standardavvikelse	t-värde	p-värde
Intercept	-0.186	0.092	-2.035	0.042
Tid	0.00008	0.00005	1.607	0.108
Månad	-0.509	0.155	-3.279	0.001
Periodicitet	-0.647	0.092	-7.021	< 0.001

Tabell 4.9 presenterar parameterskattningarna för interceptet och de tidsrelaterade förklaringsvariablerna som är skattade med modellen med strukturen ifrån ekvation 3.7.

Interceptet är signifikant skilt ifrån noll med en punktskattning på -0.186. Detta tolkas som att det finns en signifikant skillnad mellan korsningen och väglänken i början av tidsserien där den första skillnaden är skattad till -0.186. Vilket innebär att det förekommer generellt sett 0.186 mindre $PM_{1-2.5}$ i korsningen än på väglänken i början av tidsserien. Då parametrarna för *tid* och *månad* är med i modellen förväntas skillnaden mellan korsningen och väglänken ändras över tid. Variabeln *periodicitet* är signifikant med en negativ parameterskattning. Detta tolkas som att mellan 9:00 och 18:00 förekommer det generellt sett mindre $PM_{1-2.5}$ i korsningen än på väglänken än under resterande klockslag.



Figur 4.15: Skattade värden korrigerade för autokorrelation och heteroskedasticitet av $PM_{1-2.5(D)}$

Figur 4.15 visar de skattade värdena på responsvariabeln $PM_{1-2.5(D)}$ över tid som är framtagna med parameterskattningarna ifrån tabell 4.9. Figuren visar att modellen skattar både positiva och negativa värden på skillnaden av $PM_{1-2.5}$ mellan korsningen och väglänken. I april månad samt under alla dagar mellan klockan 9:00 och 18:00 är samtliga värden skattade negativa. Detta tyder på att det förekommer en generell skillnad under dessa tidpunkter.

4.2.2 Modell med förklaringsvariabler

För att analysera ifall meteorologiska, trafikrelaterade och tidsrelaterade variabler har någon påverkan på skillnaden av $PM_{1-2.5}$ mellan korsningen och väglänken studeras en multipel linjär tidsserieregression. Modellen korrigeras för heteroskedasticitet och autokorrelation igenom GARCH($s = 1, r = 0$) och Cochrane-Orcutt där $p = 5$. Strukturen på s, r och p är dessamma som den modell som presenteras i kapitel 4.2.1. Detta kommer sig är för att samma mönster i residualerna förekommer.

De förklaringsvariabler som får möjlighet till att vara med i modellen är de som presenteras i tabell 2.2.

För att avgöra vilka förklaringsvariabler som signifikant påverkar responsvariabeln tillämpas bakåteliminering och ett partiellt F-test.

Tabell 4.10: Bakåteliminering för $PM_{1-2.5(D)}$

Variabel	P-värden		
	Modell 1	Modell 2	Modell 3
Tid	0.263	0.262	-
Månad	0.013	0.013	0.007
Periodicitet	< 0.001	< 0.001	< 0.001
Relativ fuktighet	< 0.001	< 0.001	< 0.001
Maxvind	< 0.001	< 0.001	< 0.001
Temperatur	0.011	0.011	0.006
Lufttryck	0.065	0.067	0.046
Trafik	0.019	0.018	0.019
Nederbörd	0.719	-	-

Tabell 4.10 visar att variablerna *temperatur*, *lufttryck* och *nerbörd* inte bör vara med och förklara responsvariabeln $PM_{1-2.5(D)}$. Modellen reduceras därmed igenom att ta bort dessa förklaringsvariabler ifrån modellen.

För att avgöra ifall vindriktning ska vara med i modellen genomförs ett partiellt F-

test då vindriktning förklaras av tre stycken variabler.

Tabell 4.11: Partiellt F-test av vindriktning för $PM_{1-2.5(D)}$

Partiellt F-test					
Variabel	SSE	df	df residualer	F-värde	p-värde
Modell utan vindriktning	18501.6	7	5794		
Modell med vindriktning	17255.3	10	5791		
	1246.319	3		139.42	< 0.001

F-testet i tabell 4.11 visar att de tre variablerna som förklarar vindriktning bör vara med i modellen då p-värdet understiger signifikansnivån på fem procent.

När alla förklaringsvariabler som bör vara med i modellen enligt bakåteliminering och det partiella F-testet konstaterats skattas modellens parametrar.

Parametrarna som modellerar för heteroskedasticitet och autokorrelation kan ses i bilaga A.2

Efter sex iterationer förändrades inte parameterskattningarna av ϕ_1, \dots, ϕ_5 nämnvärt. Framtagning av parametrar för förklaringsvariablerna, $\beta_0, \dots, \beta_{10}$, sker därmed med sex iterationer.

Tabell 4.12 presenterar parameterskattningarna för förklaringsvariablerna som är skattade med modellen med strukturen ifrån ekvation 3.7.

Tolkning av förklaringsvariabler givet att de andra förklaringsvariablerna hålls fixa:

- **Intercept, Månad:** Den generella skillnaden av $PM_{1-2.5}$ mellan korsningen och väglänken är skattad till -0.267 under mars månad och till -0.345 under april månad givet att förklaringsvariablerna hålls på deras medelvärden.
- **Periodicitet:** Är klockan mellan 9:00 och 18:00 tenderar det till att mindre $PM_{1-2.5}$ förekommer i korsningen än på väglänken.
- **Relativ fuktighet:** Är den relativa fuktigheten hög tenderar det till att mer $PM_{1-2.5}$ förekommer i korsningen än på väglänken.
- **Maxvind:** Är maxvinden hög tenderar det till att mer $PM_{1-2.5}$ förekommer i korsningen än på väglänken.
- **Temperatur, Lufttryck, Trafik:** Då parametrarna för temperatur, lufttryck och trafik inte är signifikanta efter att förklaringsvariablerna för vindriktning lagts till i modellen bör inte de inte tolkas.

- **Riktning på vinden:** Vindriktning longitud har blivit skattad negativ vilket kan tolkas som att blåser det östlig vind tenderar det till att mindre $PM_{1-2.5}$ förekommer i korsningen än på väglänken än om det hade varit västlig vind.

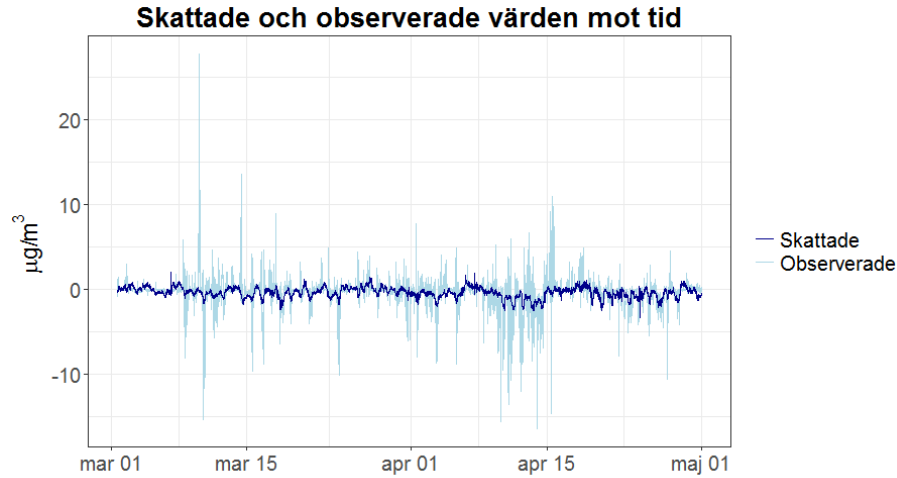
Vindriktning latitud har blivit skattad negativ vilket kan tolkas som att blåser det nordlig vind tenderar det till att mindre $PM_{1-2.5}$ förekommer i korsningen än på väglänken än om det hade varit sydlig vind.

Vindriktning interaktion är signifikant vilket tyder på att det finns en interaktion mellan vindriktning longitud och vindriktning latitud.

Tabell 4.12: Skattning av $\beta_0, \dots, \beta_{10}$ för $PM_{1-2.5}$ med förklaringsvariabler

Parameterskattningar för $PM_{1-2.5}$				
Variabel	Skattning (β_i)	Standardavvikelse	t-värde	p-värde
Intercept	-0.267	0.061	-4.381	< 0.001
Månad	-0.078	0.079	-0.985	0.325
Periodicitet	-0.370	0.099	-3.715	< 0.001
Relativ fuktighet	0.251	0.046	5.486	< 0.001
Maxvind	0.252	0.026	9.842	< 0.001
Trafik	-0.059	0.033	-1.816	0.069
Temperatur	-0.048	0.049	-0.977	0.329
Lufttryck	-0.027	0.037	-0.741	0.459
Vindriktning longitud	-0.322	0.029	-11.169	< 0.001
Vindriktning latitud	-0.282	0.031	-9.228	< 0.001
Vindriktning interaktion	-0.152	0.022	-7.056	< 0.001

Förklaringsvariabler som är standardiserade och har signifikanta parametrar är *relativ fuktighet*, *maxvind*, *vindriktning longitud*, *vindriktning latitud* och *vindriktning interaktion*. Variablerna är standardiserade och deras effekt på responsvariabeln $PM_{1-2.5}$ kan jämföras igenom att studera storleken på parameterskattningen. Förklaringsvariablerna *relativ fuktighet*, *maxvind*, *vindriktning longitud* och *vindriktning latitud* har fått en parameterskattning mellan absolutvärdena 0.25 och 0.33.



Figur 4.16: Skattade värden korrigerade för autokorrelation och heteroskedasticitet av $PM_{1-2.5}(D)$

I figur 4.16 ses de skattade och observerade värdena för responsvariabeln $PM_{1-2.5}$. Den skattade modellen fångar upp den generella trenden i responsvariabeln då de skattade värdena följer de observerade värdena relativt bra. Modellen har dock svårt att fånga upp extremvärden.

Förklaringsgraden, R^2 , för modellen är 15.44 procent.

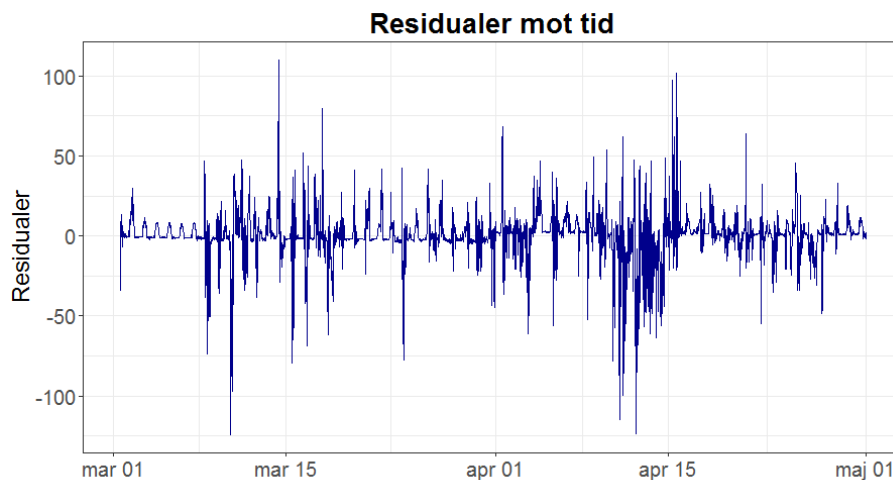
4.3 Analys av $PM_{2.5-10}(D)$

Responsvariabeln $PM_{2.5-10}(D)$ analyseras genom två separata multipla linjära tidsseriereggressionsmodeller. Den första modellen undersöker ifall det förekommer någon signifikant skillnad av $PM_{2.5-10}$ över tid mellan korsningen och väglänken. Den andra modellen undersöker ifall meteorologiska, trafikrelaterade och tidsrelaterade variabler har någon påverkan på skillnaden av $PM_{2.5-10}$ mellan korsningen och väglänken.

4.3.1 Modell med tidsvariabler

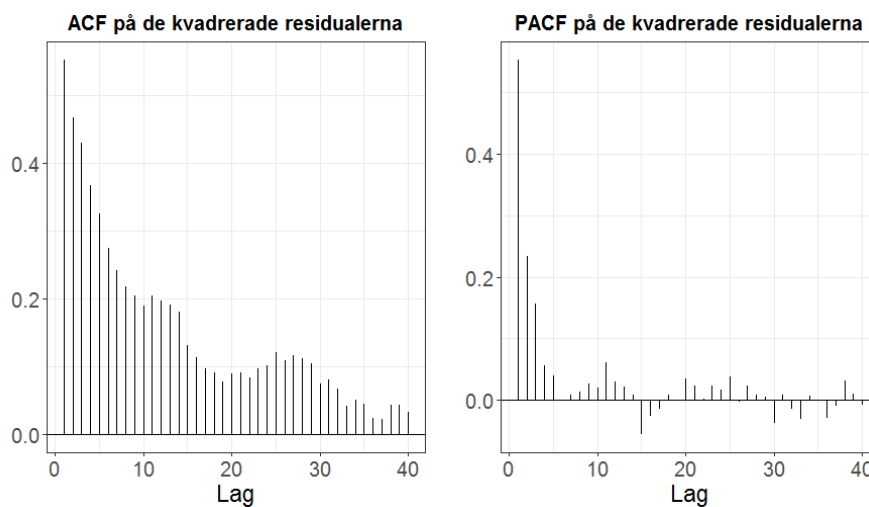
För att analysera ifall det förekommer någon signifikant skillnad av $PM_{2.5-10}$ över tid mellan korsningen och väglänken analyseras en multipel linjär tidsseriereggression med strukturen ifrån ekvation 3.1. I modellen förekommer endast tidsrelaterade variabler förklaringsvariabler så som *tid*, *månad* och *periodicitet* som presenteras i tabell 2.2.

För att undersöka ifall modellens krav, som presenteras i kapitel 3.1, är uppfylla studeras modellens residualer.



Figur 4.17: Residualer över tid för $PM_{2.5-10(D)}$

Figur 4.17 visar på att residualerna ifrån modellen med strukturen ifrån ekvation 3.1 inte har en konstant varians över tid. För att modellera för en icke konstant varians byggs en $GARCH(s,r)$ vidare på modellen. För att avgöra ordningen s och r studeras ACF och PACF på de kvadrerade residualerna.



Figur 4.18: ACF och PACF för fastställande av GARCH för $PM_{2.5-10(D)}$

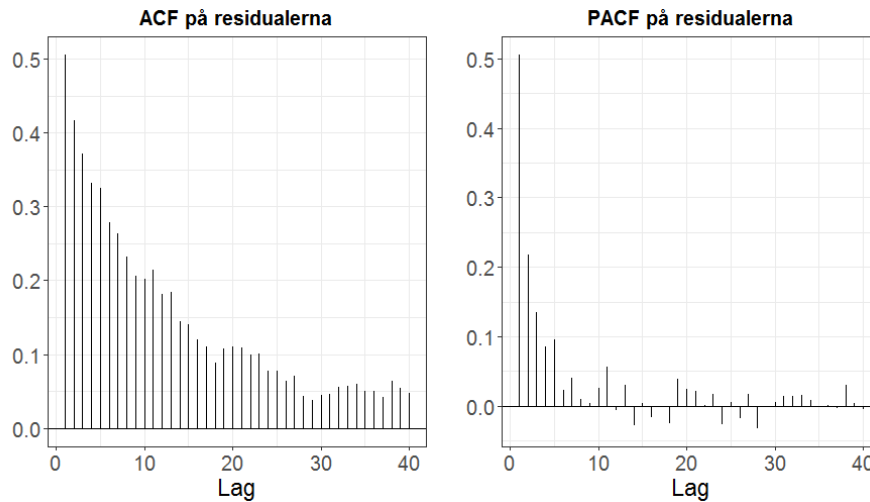
Figur 4.18 visar ett avtagande möster i ACF och tre stycken spikar konstateras vara tydliga i PACF. Ordning s utses till 3 och ordningen r utses till 0. En modell med strukturen i ekvation 3.5 kan därmed skattas.

Tabell 4.13: Skattning av α_0 , och α_1 för $PM_{2.5-10(D)}$ med tidsvariabler

Modellering för heteroskedasticitet				
Parameter	Skattning	Standardavvikelse	t-värde	p-värde
α_0	62.466	8.919	7.004	< 0.001
α_1	0.387	0.013	29.859	< 0.001
α_2	0.167	0.014	12.135	< 0.001
α_3	0.157	0.013	12.069	< 0.001

Tabell 4.13 visar de parametrar som hjälper till att modellera för heteroskedasticitet. Parametrarna skattas med hjälp av ekvation 3.4.

För att avgöra ifall modellen, som är korrigerad för heteroskedasticitet, uppfyller kraven studeras dess residualer som visas i figur 4.19.



Figur 4.19: ACF och PACF för fastställande av Cochrane-Orcutt för $PM_{2.5-10(D)}$

Figur 4.19 visar att residualerna ifrån modellen med strukturen ifrån ekvation 3.5 innehåller en positiv autokorrelation, vilket ska korrigeras för. Ett avtagande mönster ses i ACF och fem tydliga spikar ses i PACF. Modellen ska därför modelleras med en autoregressiv struktur, vilket görs med Cochrane-Orcutt som presenteras i ekvation 3.7. Fem stycken spikar konstateras vara tydliga i PACF och p bestäms till 5.

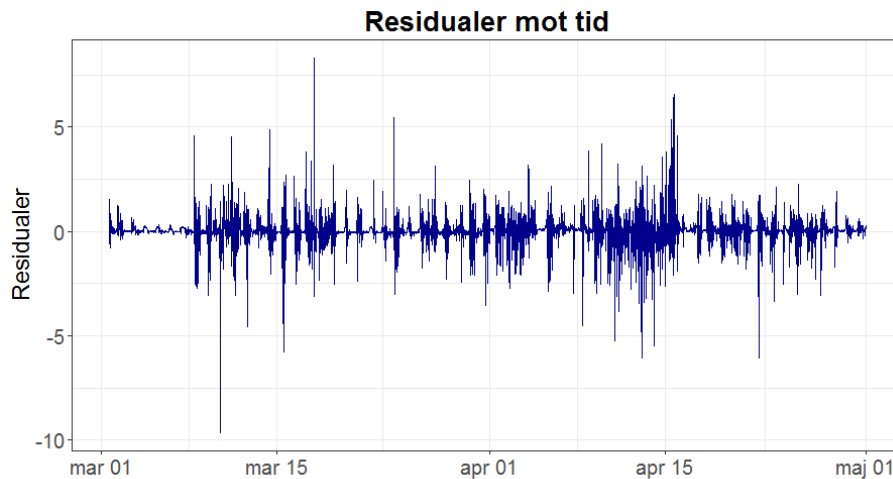
Tabell 4.14: Skattning av ϕ_1, \dots, ϕ_5 för $\text{PM}_{2.5-10(D)}$ med tidsvariabler

Modellering för autokorrelation				
Parameter	Skattning	Standardavvikelse	t-värde	p-värde
ϕ_1	0.349	0.013	26.686	< 0.001
ϕ_2	0.141	0.014	10.174	< 0.001
ϕ_3	0.092	0.014	6.596	< 0.001
ϕ_4	0.052	0.014	3.777	< 0.001
ϕ_5	0.098	0.013	7.484	< 0.001

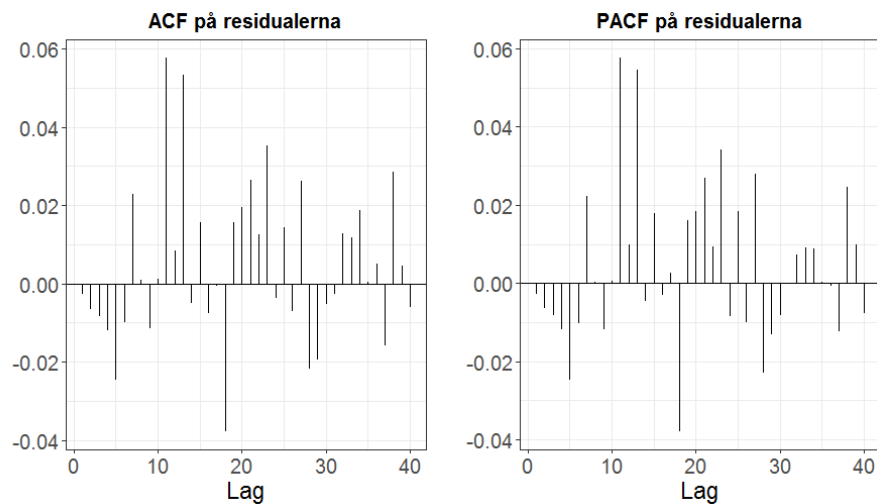
Tabell 4.14 visar parametrarna som hjälper till att modellera för autokorrelation. Parametrarna skattas med hjälp av ekvation 3.6.

Efter sju iterationer förändras inte parameterskattningarna av ϕ_1, \dots, ϕ_5 nämnvärt. Framtagning av parametrar för de tidsrelaterade förklaringsvariablerna, β_0, \dots, β_3 , sker därmed med sju iterationer.

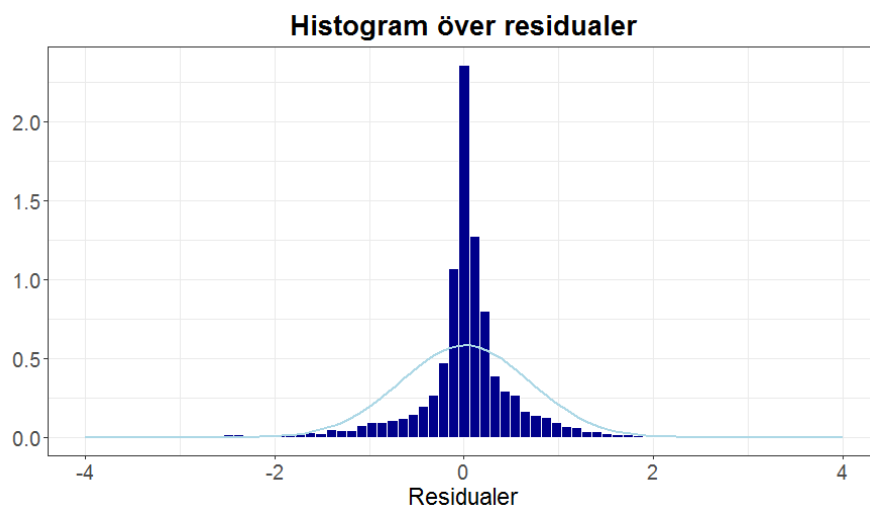
Innan β_0, \dots, β_3 studeras, studeras modellens residualer för att se ifall modellen uppfyller dess krav.



Figur 4.20: Residualer över tid korrigerade för autokorrelation och heteroskedasticitet av $\text{PM}_{2.5-10(D)}$



Figur 4.21: ACF och PACF korregerade för autokorrelation och heteroskedasticitet av $PM_{2.5-10(D)}$



Figur 4.22: Fördelning av residualer korregerade för autokorrelation och heteroskedasticitet av $PM_{2.5-10(D)}$

Figurerna 4.20 - 4.22 är baserade på residualer ifrån modellen med strukturen i ekvation 3.7.

Figur 4.20 tyder på att residualerna har en jämnare varians och figur 4.21 visar på att residualerna är okorrelerade. Figur 4.22 visar hur residualerna följer en normalfördelning. Figuren visar att fördelningen är för toppig för att följa en normalfördelning. Dock för att gå vidare med modellen antas residualerna vara approximativt normalfördelade.

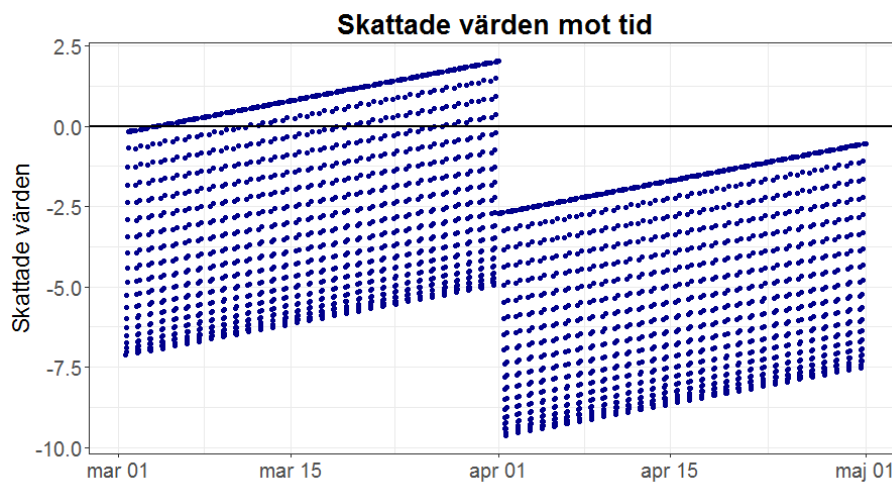
Modellen med strukturen ifrån ekvation 3.7 antas uppfylla dess krav. Denna modell är modellerad för heteroskedasticitet genom en GARCH(3,0) samt modellering för autokorrelation sker med Cochrane Orcutt där $p = 5$.

Tabell 4.15: Skattning av β_0, \dots, β_3 för $PM_{2.5-10(D)}$ med tidsvariabler

Parameterskattningar för $PM_{2.5-10(D)}$				
Variabel	Skattning (β_i)	Standardavvikelse	t-värde	p-värde
Intercept	-0.183	0.662	-0.276	0.783
Tid	0.00076	0.00034	2.226	0.026
Månad	-4.745	1.144	-4.148	< 0.001
Periodicitet	-6.957	0.764	-9.102	< 0.001

Tabell 4.15 presenterar parameterskattningar för interceptet och de tidsrelaterade förklaringsvariablerna som är skattade med modellen med strukturen ifrån ekvation 3.7.

Interceptet är inte signifikant skilt ifrån noll då p-värdet överstiger signifikansnivån på fem procent. Det finns inga belägg för att korsningens $PM_{2.5-10}$ halt skiljer sig ifrån väglänkens $PM_{2.5-10}$ i början av tidsserien. Då variablerna *tid*, *månad* och *periodicitet* är med i modellen förväntas skillnaden mellan korsningen och väglänken ändras över tid.



Figur 4.23: Skattade värden korrigerade för autokorrelation och heteroskedasticitet av $PM_{2.5-10(D)}$

Figur 4.23 visar de skattade värdena på responsvariabeln $PM_{2.5-10(D)}$ över tid som är

framtagna med parameterskattningarna ifrån tabell 4.15. Figuren visar att modellen skattar både positiva och negativa värden på skillnaden av $PM_{2.5-10}$ mellan korsningen och väglänken. I april månad samt under alla dagar mellan klockan 11:00 och 16:00 är samtliga värden skattade negativa. Detta tyder på att det förekommer en generell skillnad under dessa tidpunkter.

4.3.2 Modell med förklaringsvariabler

För att analysera ifall meteorologiska, trafikrelaterade och tidsrelaterade variabler har någon påverkan på skillnaden av $PM_{2.5-10}$ mellan korsningen och väglänken studeras en multipel linjär tidsserieregression. Modellen korrigeras för heteroskedasticitet och autokorrelation igenom GARCH($s = 3$, $r = 0$) och Cochrane-Orcutt där $p = 5$. Strukturen på s , r och p är dessamma som den modell som presenteras i kapitel 4.3.1. Detta kommer sig är för att samma mönster i residualerna förekommer.

De förklaringsvariabler som får möjlighet till att vara med i modellen är de som presenteras i tabell 2.2.

För att avgöra vilka förklaringsvariabler som signifikant påverkar responsvariabeln tillämpas bakåteliminering och ett partiellt F-test.

Tabell 4.16: Bakåteliminering för $PM_{2.5-10(D)}$

P-värden					
Variabel	Modell 1	Modell 2	Modell 3	Modell 4	Modell 5
Tid	0.0287	0.028	0.0277	0.0214	0.0084
Månad	0.0042	0.0043	0.0049	0.0055	0.0053
Periodicitet	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001
Relativ fuktighet	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001
Maxvind	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001
Temperatur	0.2543	0.2635	0.2236	0.1892	-
Lufttryck	0.2856	0.2861	0.2764	-	-
Trafik	0.4772	0.4705	-	-	-
Nederbörd	0.9201	-	-	-	-

Tabell 4.16 visar att variablerna *temperatur*, *lufttryck*, *trafik* och *nerbörd* inte bör vara med och förklara responsvariabeln $PM_{2.5-10(D)}$. Modellen reduceras därmed igenom att ta bort dessa förklaringsvariabler från modellen.

För att avgöra ifall vindriktning ska vara med i modellen genomförs ett partiellt F-

test då vindriktning förklaras av tre stycken variabler.

Tabell 4.17: Partiellt F-test av vindriktning för $PM_{2.5-10(D)}$

Partiellt F-test					
Variabel	SSE	df	df residualer	F-värde	p-värde
Modell utan vindriktning	1159090	5	5796		
Modell med vindriktning	1091224	8	5793		
	67865	3		120.09	< 0.001

F-testet i tabell 4.17 visar att de tre variablerna som förklarar vindriktning bör vara med i modellen då p-värdet understiger signifikansnivån på fem procent.

När alla förklaringsvariabler som bör vara med i modellen enligt bakåteliminering och det partiella F-testet konstaterats skattas modellens parametrar.

Parametrarna som modellerar för heteroskedasticitet och autokorrelation kan ses i bilaga A.3

Tabell 4.18: Skattning av β_0, \dots, β_8 för $PM_{2.5-10(D)}$ med förklaringsvariabler

Parameterskattningar för $PM_{2.5-10(D)}$				
Variabel	Skattning (β_i)	Standardavvikelse	t-värde	p-värde
Intercept	-2.284	0.655	-3.488	< 0.001
Tid	0.00083	0.00032	2.593	0.010
Månad	-2.774	1.081	-2.565	0.010
Periodicitet	-4.719	0.736	-6.409	< 0.001
Relativ fuktighet	3.439	0.302	11.398	< 0.001
Maxvind	1.277	0.182	7.008	< 0.001
Vindriktning longitud	-2.257	0.221	-10.196	< 0.001
Vindriktning latitud	-1.537	0.230	-6.682	< 0.001
Vindriktning interaktion	-1.048	0.157	-6.693	< 0.001

Tabell 4.18 presenterar parameterskattningarna för förklaringsvariablerna som är skattade med modellen med strukturen ifrån ekvation 3.7.

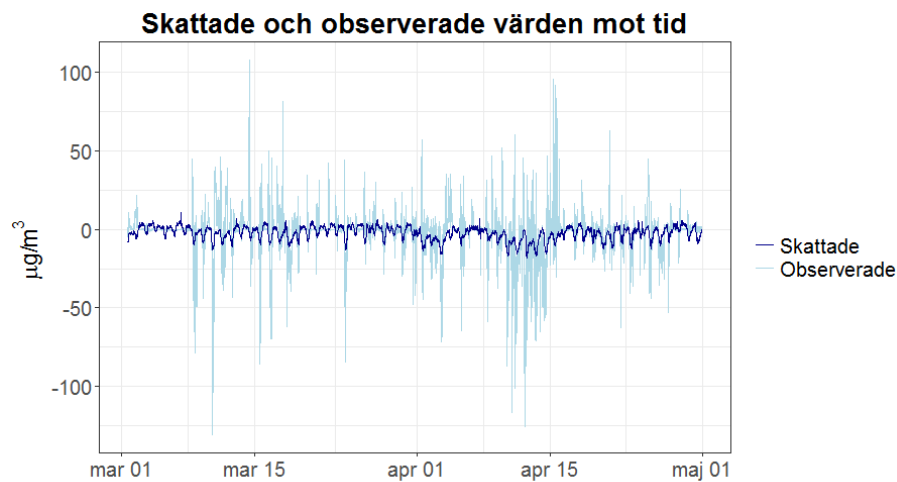
Tolkning av förklaringsvariabler givet att de andra förklaringsvariablerna hålls fixa:

- **Intercept, Tid, Månad:** Den generella skillnaden av $PM_{2.5-10}$ mellan korsningen och väglänken förändras över tid givet att de andra förklaringsvariabler hålls på dess medelvärde.
- **Periodicitet:** Är klockan mellan 9:00 och 18:00 tenderar det till att mindre $PM_{2.5-10}$ förekommer i korsningen än på väglänken.
- **Relativ fuktighet:** Är den relativa fuktigheten hög tenderar det till att mer $PM_{2.5-10}$ förekommer i korsningen än på väglänken.
- **Maxvind:** Är maxvinden hög tenderar det till att mer $PM_{2.5-10}$ förekommer i korsningen än på väglänken.
- **Riktning på vinden:** Vindriktning longitud har blivit skattad negativ vilket kan tolkas som att är det östlig vind tenderar det till att mindre $PM_{2.5-10}$ förekommer i korsningen än på väglänken än om det hade varit västlig vind.

Vindriktning latitud har blivit skattad negativ vilket kan tolkas som att är det nordlig vind tenderar det till att mindre $PM_{2.5-10}$ förekommer i korsningen än på väglänken än om det hade varit sydlig vind.

Vindriktning interaktion är signifikant vilket tyder på att det finns en interaktion mellan vindriktning longitud och vindriktning latitud.

Förklaringsvariabler som är standardiserade är relativ fuktighet, maxvind, vindriktning longitud, vindriktning latitud och vindriktning interaktion. Då dessa variabler är standardiserade kan deras effekt på responsvariabeln $PM_{2.5-10}$ jämföras igenom att studera storleken på parameterskattningen. Förklaringsvariabeln relativ fuktighet har fått den största parameterskattningen på 3.439 vilket tyder på att den har störst påverkan på responsvariabeln $PM_{2.5-10}$.



Figur 4.24: Skattade värden korregerade för autokorrelation och heteroskedasticitet av $PM_{2.5-10(D)}$

I figur 4.24 ses de skattade och observerade värdena för responsvariabeln $PM_{2.5-10(D)}$. Den skattade modellen fångar upp den generella trenden i responsvariabeln då de skattade värdena följer de observerade värdena relativt bra. Modellen har dock svårt att fånga upp extremvärden vilket kan ses som ett problem då det förekommer många extremvärden i responsvariabeln.

Förklaringsgraden, R^2 , för modellen är 14.26 procent.

5. Diskussion

5.1 Resultatdiskussion

För PM_1 , $PM_{1-2.5}$ och $PM_{2.5-10}$, där bara tidsvariabler finns med i modellerna, påvisas interceptet vara signifikant. För $PM_{2.5-10}$ finns ingen signifikant skillnad, mellan korsningen och väglänken, i början av tidsserien.

Viken månad det är påverkar skillnaden i partikelhalt, vilket kan förklaras av studien Johansson, Norman och Gustafsson (2008) utför. I studien framgår det att större partikelmängder råder under tidig vår och låga mängder mot maj. För samtliga partikelmått påverkar månaden skillnaden förklarade med tidsrelaterade variabler.

Tiderna mellan 09:00-18:00 har en signifikant påverkan för $PM_{1-2.5}$ och $PM_{2.5-10}$. Mellan dessa tider är det förmodligen en kombination med olika händelser som påverkar skillnaden. Enligt Janhäll (2005) kan partikelmått som PM_1 färdas långa sträckor och har förmodligen svårare att påverkas av lokala effekter och kan därför vara anledningen till att något mönster för periodiciteten inte finns i figur 2.2.

Vid varje modell där signifikanta förklarandevariabler finns med i modellen så påverkar dessa förklarandevariablerna skillnaden mellan korsningen och väglänken.

Det påvisas att relativa fuktigheten, maxvinden och vindriktningen signifikant påverkar skillnaden mellan korsningen och väglänken. Det är rimligt att dessa variabler påverkar skillnaden då Stockholms Luft- och Bulleranalys (SLB) i studien som Gustafsson m. fl. (2016) tidigare har konstaterat att bland annat relativa fuktighet och vindhastighet påverkar partikelhalten och därmed skulle kunna hjälpa till att förklara skillnaden av partikelhalten mellan en korsning och en väglänk.

Skillnaden mellan korsningen och väglänken kan bero på att korsningen ligger mer öppet än väglänken och påverkas av förklarande variablerna på olika sätt. Väglänken ligger positionerad där två hus ligger på vardera sida av vägen och kommer rimligtvis påverka hur vindriktning påverkar skillnaden. Korsningen ligger mer öppet då Hornsgatan korsas av en annan gata och därmed mindre hus som kan påverka hur vindriktningen påverkar skillnaden. Vindriktning har också en påverkan på den generella skillnaden, vilket också F-testet signifikant visar. Vindriktning i longitud för alla modeller är negativt skattade vilket säger att vind som kommer ifrån väst gör att partikelhalten tenderar att vara mindre i korsningen än i väglänken. Vindriktning i latitud för alla modeller är negativt skattade och signifikanta, förutom för PM_1 . Den negativa skattningen säger att när vinden kommer ifrån söder betyder det att partikelhalten tenderar att vara mindre i korsningen än i väglänken. Påverkan av skillnaden mellan korsningen och väglänken kan bero på att korsningen, som tidigare nämnt i diskussionen, ligger mer öppet och ger fler möjligheter för vinden att påverka skillnaden. Vindhastigheten har generellt också en signifikant påverkan på partikelhalterna, vilket också kan förklara varför det finns en skillnad mellan korsningen och väglänken.

Den generella trafikmängden för PM_1 och $PM_{1-2.5}$ tenderar till skillnaden mellan korsningen och väglänken. Det är förmodligen för att de mindre partiklarna, PM_1 och $PM_{1-2.5}$, påverkas mer av emissioner av fordon och finare partiklar. För $PM_{2.5-10}$ så påverkar inte den generella trafikmängden skillnaden. Det kan bero på att det blir mer uppblandat i luften på grund av genomströmningen av fordon och damm och andra partiklar, som finns på vägbanan, finns mer utspritt mellan korsningen och väglänken.

5.2 Metoddiskussion

Modellerna i uppsatsen är baserade på en felterm som är heteroskedastisk och är korri-gerade för autokorrelation. På grund av den starka ickekonstanta variansen i tidsserierna och ett stort antal extremvärden, som visas i figur 2.1, har anpassningen av modellerna behövt ta hänsyn till flera steg av residualanalys. Den ickekonstanta variansen i data är anledningen till modelleringen för heteroskedasticitet i feltermen. Modelleringen har i huvudsakligt syfte att påverka skattningarna av parametrarna så skattningen av feltermen blir mer nerskalad och det inte förekommer lika mycket extremvärden.

Figur 4.1, 4.9, 4.17 visar residualerna för respektive modell och att många värden ligger kring noll samt att många extremvärden förekommer. Ett antal värden är skattade för lågt eller för högt, vilket resulterar i de toppiga fördelningarna med smala svansar i histogrammen. Detta är för att det finns många extremvärden för alla partikelhalter, vilket visas i figur 2.1, som påverkar anpassningen av modellen och det blir svårt att skatta plötsligt förekommande extremvärden.

Förklaringsgraden i modellerna, som är 22.35, 15.44 och 14.26 procent för PM_1 , $PM_{1-2.5}$ respektive $PM_{2.5-10}$, är inte så höga och det beror på ett stort antal extremvärden som är svåra att skatta för modellerna. För att kunna skatta modellera bättre skulle tiden för dessa extremvärden undersökas där mer information om hur modellerna skulle kunna förbättras och därmed få en säkrare slutsats. En bättre skattning skulle leda till bättre anpassning till data och skulle ge högre förklaringsgrad. Det är viktigt påpeka att om dessa extremvärden inte har ett återkommande mönster och man modellerar för dessa, ska inte prediktion utföras då modellerna är överanpassade till data. Prediktion är inte heller uppsatsen syfte så tolkning för modellerna fungerar endast för det data och tidsintervallet som erhålls.

Att kolla på meteorologiska variabler för att undersöka skillnaden kanske, på grund av den låga förklaringsgraden, inte är så bra eller att någon annan variabel behövs för att komplettera modellerna. Förklaringsgraden är högre för de mindre partiklarna och lägre för de större partiklarna. Detta betyder att modellen fångar upp större andel varians och det kan bero på att extremvärden inte är lika vanligt förekommande i data för PM_1 , jämförelse med $PM_{1-2.5}$ och $PM_{2.5-10}$.

6. Slutsats

I detta kapitel kopplas frågeställningarna som presenteras i kapitel 1.4.1 ihop med de resultat som presenteras i kapitel 4.

6.1 Finns det någon signifikant skillnad över tid i halter av olika partiklar mellan korsningen och väglänken?

Resultat visar på att det finns en generell skillnad av PM_1 mellan korsningen och väglänken över hela tidsserien, där korsningen har lägre halt av PM_1 . Den förväntade skillnaden av PM_1 mellan korsningen och väglänken varierar över tid. Dock så förväntas korsningen ha en lägre generell PM_1 halt under hela tidsserien.

För $PM_{1-2.5}$ visar resultat att det finns en generell skillnad av $PM_{1-2.5}$ mellan korsningen och väglänken den större delen av tidsserien där korsningen har en lägre halt. Tidpunkter som tyder på att en generell skillnad är april månad samt alla dagar mellan klockan 9:00 och 18:00 oavsett månad.

Liknande slutsats som dras för $PM_{1-2.5}$ kan dras för $PM_{2.5-10}$. Resultat tyder på att en generell skillnad av $PM_{2.5-10}$ mellan korsningen och väglänken förekommer i april månad samt under alla dagar mellan klockan 11:00 och 16:00. Det är korsningen som tyder på att ha en lägre halt av $PM_{2.5-10}$ än väglänken.

6.2 Har meteorologiska, trafikrelaterade och tidsrelaterade variabler någon påverkan på skillnader i partikelhalter mellan korsningen och väglänken?

Förklaringsvariabler som får chansen att förklara skillnaden av de olika partikelhalterna är de som presenteras i tabell 2.2. Av dessa har inte alla förklaringsvariabler visat sig påverka skillnaden av partikelhalterna mellan korsningen och väglänken.

Tabell 6.1 visar vilka förklaringsvariabler som signifikant påverkar skillnaden av respektive partikelhalt mellan korsningen och väglänken. Förklaringsvariabler som påverkar samtliga skillnader är *relativ fuktighet*, *maxvind* och *vindriktning* i framförallt *longitud riktning*.

I följande punkter tolkas de meteorologiska, trafikrelaterade och tidsrelaterade variablernas påverkan på skillnaden i partikelhalt mellan korsningen och väglänken. Där inte storlek på partikelhalt nämns är det samma tolkning för de tre partikelhalterna.

- **Tid, Månad:** Förklaringsvariablerna tid och månad är med i modellen för att

förklara hur den generella skillnaden av partikelhalterna har förändrats över tid.

- **Periodicitet:** Mellan klockan 9:00 och 18:00 tenderar det till att mindre $PM_{1-2.5}$ och $PM_{2.5-10}$ förekommer i korsningen än på väglänken.
- **Relativ fuktighet:** Är den relativa fuktigheten hög tenderar det till att mer $PM_{1-2.5}$ och $PM_{2.5-10}$ förekommer i korsningen än på väglänken. Däremot då den relativa fuktigheten är hög tenderar det till att mindre PM_1 förekommer i korsningen än på väglänken.
- **Maxvind:** Är maxvinden hög tenderar det till att mer partikelhalt förekommer i korsningen än på väglänken.
- **Trafik:** Är det mycket trafik tenderar det till att mindre PM_1 förekommer i korsningen än på väglänken.
- **Riktning på vinden:** Är det östlig vind tenderar det till att mindre partikelhalt förekommer i korsningen än på väglänken än om det hade varit västlig vind.

För $PM_{1-2.5}$ och $PM_{2.5-10}$ gäller även att om det är nordlig vind tenderar det till att mindre partikelhalt förekommer i korsningen än på väglänken än om det hade varit sydlig vind.

Det förekommer även en interaktion mellan vindriktning longitud och vindriktning latitud.

Tabell 6.1: Partikelhalternas påverkan av förklaringsvariabler

Förklaringsvariabel	$PM_{1(D)}$	$PM_{1-2.5(D)}$	$PM_{2.5-10(D)}$
Tid	X		X
Månad	X		X
Periodicitet		X	X
Relativ fuktighet	X	X	X
Maxvind	X	X	X
Trafik	X		
Temperatur			
Lufttryck			
Nederbörd			
Vindriktning longitud	X	X	X
Vindriktning latitud		X	X
Vindriktning interaktion	X	X	X

Litteratur

Bowerman, Bruce L., Richard T. O'Connell och Anne B. Koehler. *Forecasting, Time series and Regression*. 4. utg. Curt Hinrichs, 2004, s. 195, 406–417.

Box, George E.P., Gwilym M. Jenkins, Greta M. Ljung och Gregory C. Reinsel. *Time Series Analysis: Forecasting and Control*. Fifth edit. New Jersey: John Wiley & Sons, Inc., 2016, s. 83,298–300,367. ISBN: 978-1-118-67502-1.

Camfil. *Luftburna partiklar1 – nytt fokus på människors hälsa*. 2012.

Chen, Cathy W.S., Muyi Li, Nga T.H. Nguyen och Songsak Sriboonchitta. "On Asymmetric Market Model with Heteroskedasticity and Quantile Regression". I: *Computational Economics* 49.1 (2015), s. 155–174. ISSN: 15729974. DOI: 10.1007/s10614-015-9550-3.

Cochrane, D. och G. H. Orcutt. "Application of Least Squares Regression to Relationships Containing Auto- Correlated Error Terms". I: *Journal of the American Statistical Association* 44.245 (1949), s. 32–61.

Donders, A. Rogier T., Geert J.M.G. van der Heijden, Stijnen Theo och Karel G M Moons. "Review: A gentle introduction to imputation of missing values". I: *Journal of Clinical Epidemiology* 59.10 (2006), s. 1087–1091. ISSN: 08954356. DOI: 10.1016/j.jclinepi.2006.01.014.

Europaparlamentet och Europeiska unionens råd. "Europaparlamentets and rådets direktiv 2008/50/EG". I: 38.850 (2008), s. 1–62.

Forsberg, Bertil och Bo Segerstedt. "Vägdamm and grova partiklars effekter på befolkningens hälsa". I: 2004:136 (2004), s. 1–17.

Gustafsson, Mats. "Väg- and gaturengöring som åtgärd mot höga partikelhalter orsakade av vägdamm". I: 0347-6049 (2002), s. 1–57.

Gustafsson, Mats, Michael Norman, Christer Johansson, Göran Blomqvist och Sara Janhäll. "Operational measures against PM10 pollution in Stockholm. Evaluation of Winter season 2014–2015". I: (2016), s. 1–83.

Janhäll, Sara. "-The Urban Aerosol-, Sources and Atmospheric Processes". I: 245 (2005), s. 6–7.

Johansson, Christer, Michael Norman och Mats Gustafsson. "Genomsnittliga emissionsfaktorer för PM10". I: SLB 2:2008 (2008), s. 1–23.

Miljö och Energidepartementet. *Luftkvalitetsförordning*. 2010. URL: http://www.riksdagen.se/sv/dokument-lagar/dokument/svensk-forfattningssamling/luftkvalitetsforordning-2010_sfs-2010-477 (hämtad 2017-03-02).

Samet, Jonathan M., Scott L. Zeger, Francesca Dominici, Frank Curriero, Ivan Coursac, Douglas W. Dockery, Joel Schwartz och Antonella Zanobetti. "The National Morbi-

dity, Mortality, and Air Pollution Study. Part II: Morbidity and mortality from air pollution in the United States". I: *Res Rep Health Eff Inst.* 94.Pt2 (2000), s. 79. ISSN: 1041-5505. DOI: 10.1007/s00484-008-0189-9.

VTI. *Statens väg- and transportforskningsinstitut.* 2017. URL: <https://www.vti.se/sv/forskningsomraden/> (hämtad 2017-03-08).

Wooldridge, Jeffrey M. *Introductory Econometrics.* Vol. 120-121. 2011, s. 264–265, 345–351, 377, 412–413, 417–422, 436. ISBN: 9780324581621. DOI: 10.1016/j.jconhyd.2010.08.009.

A. Tabeller

A.1 $PM_{1(D)}$

Tabell A.1: Skattning av $\alpha_0, \dots, \alpha_6$ för $PM_{1(D)}$ med förklaringsvariabler

Modellering för heteroskedasticitet				
Parameter	Skattning	Standardavvikelse	t-värde	p-värde
α_0	0.361	0.053	6.766	< 0.001
α_1	0.261	0.013	19.887	< 0.001
α_2	0.053	0.014	3.909	< 0.001
α_3	0.045	0.013	3.343	< 0.001
α_4	0.120	0.013	8.918	< 0.001
α_5	0.074	0.014	5.446	< 0.001
α_6	0.058	0.013	4.416	< 0.001

Tabell A.2: Skattning av ϕ_1, ϕ_2 och ϕ_3 för $PM_{1(D)}$ med förklaringsvariabler

Modellering för autokorrelation				
Parameter	Skattning	Standardavvikelse	t-värde	p-värde
ϕ_1	0.323	0.013	25.009	< 0.001
ϕ_2	0.223	0.013	16.768	< 0.001
ϕ_3	0.181	0.013	14.014	< 0.001

A.2 $\text{PM}_{1-2.5}(D)$

Tabell A.3: Skattning av α_0 och α_6 för $\text{PM}_{1-2.5}$ med förklaringsvariabler

Modellering för heteroskedasticitet				
Parameter	Skattning	Standardavvikelse	t-värde	p-värde
α_0	0.934	0.179	5.231	< 0.001
α_1	0.682	0.010	70.943	< 0.001

Tabell A.4: Skattning av ϕ_1, \dots, ϕ_5 för $\text{PM}_{1-2.5}$ med förklaringsvariabler

Modellering för autokorrelation				
Parameter	Skattning	Standardavvikelse	t-värde	p-värde
ϕ_1	0.197	0.013	15.075	< 0.001
ϕ_2	0.133	0.013	10.024	< 0.001
ϕ_3	0.158	0.013	11.942	< 0.001
ϕ_4	0.066	0.013	4.964	< 0.001
ϕ_5	0.103	0.013	7.892	< 0.001

A.3 $\text{PM}_{2.5-10(D)}$

Tabell A.5: Skattning av α_0 och α_3 för $\text{PM}_{2.5-10(D)}$ med förklaringsvariabler

Modellering för heteroskedasticitet				
Parameter	Skattning	Standardavvikelse	t-värde	p-värde
α_0	57.838	8.229	7.029	< 0.001
α_1	0.365	0.013	28.135	< 0.001
α_2	0.169	0.014	12.373	< 0.001
α_3	0.151	0.013	11.631	< 0.001

Tabell A.6: Skattning av ϕ_1, \dots, ϕ_5 för $\text{PM}_{2.5-10(D)}$ med förklaringsvariabler

Modellering för autokorrelation				
Parameter	Skattning	Standardavvikelse	t-värde	p-värde
ϕ_1	0.344	0.013	26.287	< 0.001
ϕ_2	0.135	0.014	9.761	< 0.001
ϕ_3	0.092	0.014	6.580	< 0.001
ϕ_4	0.041	0.014	2.964	0.003
ϕ_5	0.082	0.013	6.294	< 0.001

B. Kod i R

B.1 Modell

```
reg <- lm(Respons ~ Förklarandevariabler, data = data)
reg2 <- reg
summary_reg <- summary(reg)
summary_reg
```

B.2 ARCH

```
antal_alpha <- 6
antal_alpha <- antal_alpha+1
X <- model.matrix(reg)
Y <- model.response(model.frame (reg))

e <- (reg$residuals)^2
n <- length(e)

e_varden <- e[antal_alpha:n]
e_varden <- as.data.frame(e_varden)
for(i in 0:(antal_alpha-1)){
  e_varden[, (i+1)] <- e[(antal_alpha-i):(n-i)]
}

e_varden <- as.matrix(e_varden)
regP <- lm(e_varden[,1] ~ e_varden[,2:ncol(e_varden)])
ht <- fitted(regP)
ht_sqrt <- 1/sqrt(ht)
X <- X[antal_alpha:n,]*ht_sqrt
Y <- Y[antal_alpha:n]*ht_sqrt
reg <- lm(Y~X-1)
summary_reg <- summary(reg)
summary_reg
```

B.3 Cochrane-Orcutt

```
antal_rho<-4

skatta_orcutt_sjalv<-function(reg=reg,antal_rho=3){
  ##### X och Y #####
  X <- model.matrix(reg)
  Y <- model.response(model.frame (reg))
  ##### Skatta rho #####
  antal_rho<-antal_rho+1
  n <- length(Y)
  e <- reg$residuals

  e_varden <- e[antal_rho:n]
  e_varden <- as.data.frame(e_varden)
  for(i in 0:(antal_rho-1)){
    e_varden[, (i+1)]<-e[(antal_rho-i):(n-i)]
  }
  e_varden<-as.matrix(e_varden)
  regP <- lm(e_varden[,1] ~ e_varden[,2:ncol(e_varden)]-1)
  rho <- summary(regP)$coeff[1:(antal_rho-1)]
  rho2<-list()
  rho2[[1]] <- c(rho)

  ##### Skapa X_stjarna och Y_stjarna samt skatta modell #####
  XB <- X[antal_rho:n, ]
  YB<-Y[antal_rho:n]
  for(i in 1:(antal_rho-1)){
    XB <- XB- rho[i]*X[(antal_rho-i):(n-i),]
    YB <- YB- rho[i]*Y[(antal_rho-i):(n-i)]
  }
  regCO <- lm(YB ~ XB-1)

  ##### Berakna nya residualer med de nya beta-varдена #####
  ypCO <- as.matrix(X[,]) %*% regCO$coeff[]
  e <- Y-ypCO
  e_varden <- e[antal_rho:n]
  e_varden <- as.data.frame(e_varden)
  for(i in 0:(antal_rho-1)){
    e_varden[, (i+1)]<-e[(antal_rho-i):(n-i)]
  }
  e_varden<-as.matrix(e_varden)
```

```

##### Skatta rho #####
regP <- lm(e_varden[,1] ~ e_varden[,2:ncol(e_varden)]-1)
rho <- summary(regP)$coeff[1:(antal_rho-1)]
rho2[[2]] <- c(rho)

##### Upprepa skattningen av rho till konvergens #####
i <- 2
for(i in 2:10){
  ##### Skapa X_stjarna och Y_stjarna samt skatta modell #####
  XB <- X[antal_rho:n, ]
  YB<-Y[antal_rho:n]
  for(j in 1:(antal_rho-1)){
    XB <- XB- rho[j]*X[(antal_rho-j):(n-j),]
    YB <- YB- rho[j]*Y[(antal_rho-j):(n-j)]
  }
  regC0 <- lm(YB ~ XB-1)
  ### Berakna nya residualer med de nya B-varдена ###
  ypC0 <- as.matrix(X[,]) %*% regC0$coeff[]
  e <- Y-ypC0
  e_varden <- e[antal_rho:n]
  e_varden <- as.data.frame(e_varden)
  for(j in 0:(antal_rho-1)){
    e_varden[, (j+1)]<-e[(antal_rho-j):(n-j)]
  }
  e_varden<-as.matrix(e_varden)
  ##### Skatta rho #####
  regP <- lm(e_varden[,1] ~ e_varden[,2:ncol(e_varden)]-1)
  rho <- summary(regP)$coeff[1:(antal_rho-1)]
  # summary(regP)

  i <- i + 1
  rho2[[i]] <- rho
}
# Iteration slut #
return(regC0)
}

```