



Stockholms  
universitet

STOCKHOLMS UNIVERSITET  
Statistiska Institutionen

**Utökade användningsområden för trafikarbetets förändring**

Expanded uses for the change in traffic density

Magnus Kjellman

15-högskolepoängsuppsats inom Statistik III, ht 2012

Handledare: Mikael Möller



Stockholms  
universitet

STOCKHOLMS UNIVERSITET  
Statistiska Institutionen

## **Förord**

Grunden till denna uppsats lades under diskussioner med Trafikverket vad deras unika datamaterial skulle kunna användas till, utöver rena trafikprognoser. Det har under en längre tid funnits idéer på att undersöka sambandet mellan förändring i trafik och BNP men inget konkret arbete har i Sverige genomförts på området. Författaren vill därför framför sitt tack till Dennis Andersson med kollegor på trafikverket som med sitt entusiastiska stöd gjorde denna uppsats möjlig att genomföra.

## **Sammanfattning**

Samlat hos många statliga verk finns en stor mängd data som oftast används i ett speciellt syfte även om de potentiella användningsområdena är långt fler. Syftet med denna uppsats är att undersöka om data från Trafikverket över förändringar i trafikmängden i det svenska vägnätet kan användas för andra syften. Studien genomfördes på data från perioden januari 1992 till augusti 2011. Detta analyserades sedan med hjälp av ARIMA-modeller. Resultatet indikerar att datat går att tillämpa på alternativa användningsområden.

## **Abstract**

There are a large amount of data gathered at many state institutions. Usually this data is only used for a specific purpose even though the potential uses are far greater. The purpose of this paper is to investigate if data collected by the Swedish Transport Administration of changes in traffic density in the Swedish road network can be used for other purposes. The study uses data from January 1992 to August 2011. The data was analyzed used ARIMA-models. The result indicates that the data can be used for different applications.



<b>1 Inledning.....</b>	<b>3</b>
1.1 Bakgrund.....	3
1.2 Syfte .....	3
1.3 Avgränsning.....	3
1.4 Metod .....	3
<b>2 Metoder för tidsserieanalys.....</b>	<b>4</b>
2.1 Begrepp.....	4
2.2 ARIMA.....	4
2.2.1 AR.....	4
2.2.2 MA.....	5
2.2.3 ARMA.....	6
2.2.4 ARIMA.....	7
2.3 Modellering och utvärderingsmetoder.....	8
<b>3 Datamaterial.....</b>	<b>10</b>
3.1 Trafikdata.....	10
3.1.1 Stratifiering.....	11
3.1.2 Skattningsmodell.....	11
3.1.3 Transformation .....	12
3.2 BNP.....	12
<b>4 Resultat.....</b>	<b>13</b>
4.1 Trafik.....	13
4.2 BNP.....	14
<b>5 Slutsats.....</b>	<b>14</b>
<b>6 Diskussion.....</b>	<b>14</b>
6.1 Datamaterial.....	14
6.2 Studiens genomförande och begränsningar.....	15
6.3 Vidare analysmöjligheter.....	15
<b>7 Referenser.....</b>	<b>16</b>
<b>Appendix 1.....</b>	<b>17</b>

## 1 Inledning

### 1.1 Bakgrund

Prediktioner av BNP har sedan begreppet infördes under 1950-talets mitt (Lequillier, F. & Blades, D., 2006) varit av stort intresse. För närvarande används ett antal olika variabler när BNP predikteras som exempelvis industriell produktion, försäljning i detaljhandel samt olika typer av finansiell data (Barhoumi, K. mfl., 2008).

De senaste 40 åren har Trafikverket (tidigare Vägverket) samlat data om trafiken på svenska vägar i syfte att avgöra var underhåll av vägnätet ska koncentreras. Dessa data har sedan legat till grund för skattningar av förändringen i trafikmängden i det svenska vägnätet.

Den svenska finansministern Gunnar Sträng hävdade att ju fler vagnar det fanns på godstågen desto bättre var utsikterna för den svenska ekonomin (Bernhardsson, J., 2002). Det har under lång tid inspirerat andra till att undersöka om ett sådant samband skulle kunna bevisas och då inte bara för tågagnar utan även för olika typer av vägtrafik. Denna fråga lyftes under diskussioner om lämpligt ämne för en uppsats. Författaren har därefter i samarbete med Trafikverket formulerat ett syfte som dels tillfredsställer kraven för en akademisk kandidatuppsats men även Trafikverkets intressen.

### 1.2 Syfte

Syftet med denna uppsats är att analysera tidsserier för tung trafik och undersöka om de innehåller tillräckligt med information för att ligga till grund för sedvanliga tidsseriemodeller samt att undersöka om dessa serier samvarierar med BNP.

Denna analys ska i ett senare skede eventuellt ligga till grund för vidare analys vars mål är att fastställa om förändringar i tung trafik kan användas som en prediktor för BNP.

### 1.3 Avgränsning

Studien avgränsas till att endast undersöka likartade beteenden hos förändringen i tung trafik, total lastbilstrafik och BNP för perioden 1994 till 2011 i Sverige.

### 1.4 Metod

Tidsserierna kommer att analyseras med hjälp av ARIMA-modeller samt korrelationsstrukturer för att undersöka likartade beteenden och skillnader.

## 2 Metoder för tidsserieanalys

### 2.1 Grundläggande Begrepp och antaganden

Syftet med en tidsserieanalys är att genom studier av tidsseriedata dra slutsatser och observera olika karaktärsdrag. Dessa kan sedan användas för att skapa modeller vars syfte är att prediktera framtida värden. Den del av prediktionen som inte förklaras av modellen och kan anses vara en slumpfaktor benämns som vitt brus,  $w_t$ .

Ett grundläggande antagande vid användning av de flesta metoder för tidsserieanalys och modellering bygger på att den observerande variabeln är stationär. Med detta menas att sannolikhetsfördelning för  $y$  är konstant för all möjliga delmängder (Shumway, R. & Stoffer, D., 2006). I de flesta fall är dock denna definition av stationäritet allt för strikt, i stället används *svag stationäritet* vilket definieras givet en tidsserie  $\{x_t\}_t$  för vilket det gäller (Gujarati, D., 2004);

1) Medelvärde är konstant

$$E(X_t) = \mu \quad (2.1)$$

2) Kovariansen beror endast på tidsavståndet

$$E[(X_t - \mu)(X_{t+k} - \mu)] = f(k) \quad (2.2)$$

Då sägs tidsserien vara svagt stationär. Av definitionen följer speciellt att variansen är konstant då

$$V(X_t) = \text{Cov}(X_t, X_{t-1}) = f(0) \quad (2.3)$$

### 2.2 ARIMA

ARIMA presenterades 1970 av G.E.P. Box and G.M. Jenkins i Time series analysis: Forecasting and control och bygger på *autoregressive* (AR) och *moving average* (MA) modeller tillämpande på svagt stationära tidsserier (Montgomery mfl. 2006). Här efter avses "svag stationäritet" när jag skriver stationäritet.

#### 2.2.1 AR

Autoregressiva modeller består av en regression av  $x_t$  på  $x_{t-1}$  t.o.m.  $x_{t-p}$ , där  $p$  anger det antalet steg bakåt som krävs för att prediktera värdet för tidpunkt  $t$ . Nedan ses en generell AR( $p$ )-modell (Shumway, R. & Stoffer, D., 2006);

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_p x_{t-p} + w_t \quad (2.4)$$

där  $x_t$  är stationär,  $\phi_1, \phi_2, \dots, \phi_p$  är konstanter ( $\phi_p \neq 0$ ) och  $w_t$  är vitt brus som antas vara normalfördelad med medelvärde noll och varians  $\sigma^2$ .

STOCKHOLMS UNIVERSITET  
Statistiska Institutionen

Om  $E(x) \neq 0$  infogas konstanten  $\alpha = \mu(1 - \delta_1 - \dots - \delta_p)$  till modellen som då tar formen;

$$x_t = \alpha + \phi_1 x_{t-1} + \dots + \phi_p x_{t-p} + w_t \quad (2.5)$$

AR(p)-modellen kan då även skrivas med *bakåoperatorn* B;

$$\phi(B)x_t = \alpha + w_t \quad (2.6)$$

där B definieras av

$$Bx_t = x_{t-1} \quad (2.7)$$

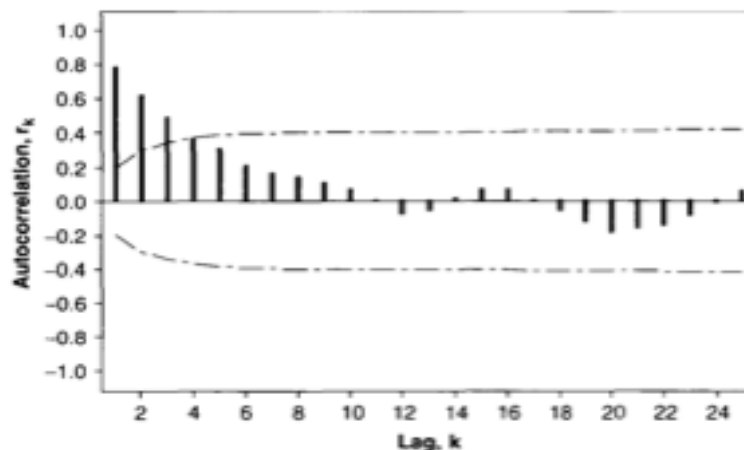
och

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p \quad (2.8)$$

För att estimer ordningen av AR-modellen används den partiella autokorrelationsfunktionen (PACF). Till skillnad från autokorrelationsfunktionen (ACF) vars skattningar inte bara inkluderar korrelationen mellan  $y_t$  och  $y_{t-k}$  utan även alla mellanliggande värden medför användandet av PACF att inflytandet från dessa mellanliggande värden ( $y_{t-1}, y_{t-2}, \dots, y_{t-k+1}$ ) rensas bort. För en AR(p)-modell ska, för  $k > p$ , PACF vara 0 (Montgomery mfl., 2006).

Autokorrelationsfunktionen kan även användas för att kontrollera om modellen ifråga är stationär (Montgomery, 2006). Detta görs genom att skatta  $\rho_k$  och göra ett diagram för  $\hat{\rho}_k, k = 1, 2, \dots$ , Figur 2.1 visar ett exempel på hur  $\hat{\rho}_k$  kan se ut för en stationär tidsserie.

Figur 2.1<sup>1</sup> Exempel på stationär tidsserie

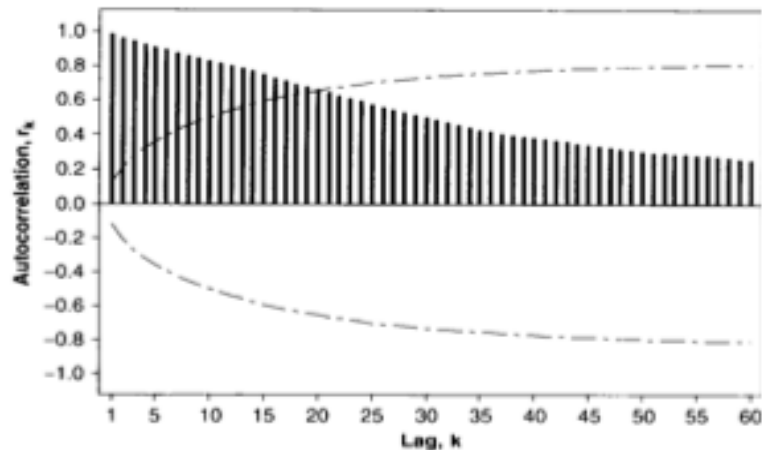


<sup>1</sup> Montgomery, 2008, s.32

STOCKHOLMS UNIVERSITET  
Statistiska Institutionen

Figur 2.2 visar hur  $\hat{\rho}$  kan bete sig när det ej kan anses föreligga stationaritet

Figur 2.2<sup>2</sup> Exempel på icke-stationär tidsserie



### 2.2.2 MA

Till skillnad från AR modeller bygger MA-modeller på antagandet att det finns ett linjärt samband mellan  $x_t$  och feltermerna  $w_{t-q}, \dots, w_t$ . En generell MA-modell av ordning  $q$ , MA( $q$ ) skrivs

$$x_t = w_t + \theta_1 w_{t-1} + \theta_2 w_{t-2} + \dots + \theta_q w_{t-q} \quad (2.9)$$

Där  $q$  är antalet steg bakåt och  $\theta_1, \theta_2, \dots, \theta_q$  ( $\theta_q \neq 0$ ) är konstanter,  $w_t$  är vitt brus som antas vara normalfördelad med medelvärde noll och konstant varians,  $\sigma^2$  (Shumway, R. & Stoffer, D., 2006).

MA( $q$ )-modellen kan även den uttryckas med hjälp av bakåtoperatören  $B$ ,

$$\theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q \quad (2.10)$$

Vid identifiering av  $q$  används autokorrelationsfunktionen (AFC). Då alla MA-modeller är stationära kommer alla autokorrelationer av större ordning än  $q$  vara noll (Kirchgassner G., Wolters J., 2007).

MA( $q$ )-modeller måste alla uppfylla kravet på *invertibilitet* vilket kortfattat innebär att de ska gå att uttrycka som en oändlig stationär AR( $p$ )-modell (Montgomery mfl., 2006).

STOCKHOLMS UNIVERSITET  
Statistiska Institutionen

### 2.2.3 ARMA

En ARMA modell är en modell som kombinerar en autoregressiv modell (AR) med en glidande medelvärdesmodell (MA). En stationär ARMA (p,q) modell kan då generellt skrivas som;

$$x_t = \alpha + \phi_1 x_{t-1} + \dots + \phi_p x_{t-p} + w_t + \theta_1 w_{t-1} + \dots + \theta_q w_{t-q} \quad (2.11)$$

där  $\phi_p \neq 0$ ,  $\theta_q \neq 0$  och  $w_t \sim ON(0, \sigma^2)$  (Shumway & Stoffer, 2006).

ARMA-modellen i (2.11) uttryckt med hjälp av bakåtoperatorer får formen;

$$\phi(B)x_t = \alpha + \theta(B)w_t \quad (2.12)$$

Det bli nu tydligt att AR- och MA-modeller är specialfall av en ARMA-modell som saknar antingen AR-komponent, dvs  $p = 0$  eller MA-komponent, dvs  $q = 0$ .

### 2.2.4 ARIMA

I processerna som beskrivits ovan har det antagits att den aktuella tidsserien är stationär. Allt som oftast är dock så inte fallet, de flesta tidsserier uppvisar någon form av trend även om beteendet är homogent över tid. För att hantera dessa icke-stationära processer används ofta *differenser*, exempel på en differens av första ordningen är;

$$(1 - B)x_t = x_t - x_{t-1} \quad (2.13)$$

Om en tidsserie blir stationär, och kan beskrivas med en ARMA-modell, genom differentiering av ordning  $d$ , dvs;

$$Y_t = (1 - B)^d X_t \quad (2.14)$$

kan den benämnas den *integrated autoregressive moving average model* (J. D. Cryer & Kung-Sik Chan, 2008). En generell ARIMA-modell, ARIMA(p,d,q) där  $d$  anger antalet differenser, skrivs med hjälp av bakåtoperatorn (Montgomery, 2006);

$$\phi(B)(1 - B)^d x_t = \alpha + \theta(B)w_t \quad (2.15)$$

Den enklaste formen av ARIMA-modell är en s.k. vittbrusprocess, ARIMA(0,0,0),

$$y_t = w_t \quad (2.16)$$



STOCKHOLMS UNIVERSITET  
Statistiska Institutionen

Många ekonomiska tidsserier, som BNP eller försäljning, uppvisar ofta säsongliknande beteende. BNP mäts ofta kvartalsvis och en period, i detta fall ett år, utgörs då av maximal 4 säsonger. Om data istället mäts på månadsbasis består en period av upp till 12 säsonger. I dessa situationer kan det vara lämpligt att använda en säsongskomponent till den aktuella modellen. En generell sådan för ett stationärt säsongsbeteende kan skrivas som, ARMA(P,Q)<sub>s</sub> ;

$$\Phi_p(B^s)x_t = \Theta_Q(B^s)w_t \quad (2.17)$$

där 
$$\Phi_p(B^s) = 1 - \Phi_1 B^s - \Phi_2 B^{2s} - \dots - \Phi_p B^{Ps} \quad (2.18)$$

och 
$$\Theta_Q(B^s) = 1 + \Theta_1 B^s + \Theta_2 B^{2s} + \dots + \Theta_Q B^{Qs} \quad (2.19)$$

(2.17) och (2.18) är en autoregressiv- respektive en glidande medelvärdes-modell för säsongsbeteende av ordning  $P$  och  $Q$  med en säsongperiod av längden  $s$  (Shumway, R. & Soffer, D., 2006).

En generell ARIMA-modell som innehåller komponenter för både säsong- och icke-säsongsbeteende, ARIMA(p,d,q)(P,D,Q), kan då skrivas som;

$$\alpha + \Phi_p(B^s)\phi(B)(1-B)^d(1-B^s)^D x_t = \alpha + \Theta_Q(B^s)\theta(B)w_t \quad (2.20)$$

Parametrarna för säsongsdelen identifieras liksom parametrarna i icke-säsongsdelen med hjälp av ACF och PACF. En ACF som uppvisar signifikanta värden på tidsavstånden 12, 24 och 36 indikerar att  $s=12$  och  $P=3$ .

**Tabell 2.1. Beteendet hos ACF och PACF för ARMA-modeller med endast säsongskomponent**

	AR(P) <sub>s</sub>	MA(Q) <sub>s</sub>	ARMA(P,Q) <sub>s</sub>
ACF*	Avtar på tidsavstånd $ks$ $k = 1, 2, \dots$	Upphör efter tidsavstånd $Qs$	Avtar på tidsavstånd $ks$
PACF*	Upphör efter tidsavstånd $Ps$	Avtar på tidsavstånd $ks$ $k = 1, 2, \dots,$	Avtar på tidsavstånd $ks$

\*Värdena för icke-säsongsfördröjningar  $h \neq ks$  för  $k = 1, 2, \dots$ , är noll. (Shumway, R. & Soffer, D., 2006)

STOCKHOLMS UNIVERSITET  
Statistiska Institutionen

### 2.3 Modellering och utvärderingsmetoder

Som tidigare nämnts kan ordningen för AR- och MA-komponenterna i en ARIMA-modell identifieras med hjälp av ACF och PACF. I 2.2.4 beskrivs även hur icke-stationära data behandlas. Att identifiera exakt av vilken ordning de olika komponenterna är av kan dock ibland vara problematiskt. För att avgöra vilken modell som passar datamaterialet bäst är det alltid viktigt att spara ett antal observationer som kan användas för att utvärdera modellens precision.

För att ytterligare utvärdera olika modellers prestation går det att använda diverse *informationskriterium*. Två sådana är *Akaiques Informationskriterium* (AIC) och *Schwarz Informationskriterium*<sup>3</sup> (SIC). Dessa är båda baserade på analys av residualer men skiljer sig något åt (Montgomery, 2006).

Schwarz Informationskriterium skrivs som (Shumway, R. & Stoffer, D., 2006);

$$SIC = \ln \hat{\sigma}_k^2 + \frac{k \ln(n)}{n} \quad (2.21)$$

där  $k$  är antal parametrar i modellen,  $n$  är urvalets storlek och

$$\hat{\sigma}_k^2 = \frac{RSS_k}{n} \quad (2.22)$$

där  $RSS_k$  är den kvadrerade summan av residualerna från en modell med  $k$  parametrar.

Akaiques Informationskriterium skrivs som;

$$AIC = \ln \hat{\sigma}_k^2 + \frac{n+2k}{n} \quad (2.23)$$

där  $\hat{\sigma}_k^2$  ges av (2.22),  $k$  är antalet parametrar i modellen och  $n$  är urvalets storlek.

Tanken är att en modell med låg varians är att föredra. En nackdel med AIC är att AIC misslyckas med att korrigera för lägre varians som uppstår på grund stort antal parametrar. För att lösa den problematiken kan istället en korrigerad variant av AIC, AICc, användas;

$$AICc = \ln \hat{\sigma}_k^2 + \frac{n+k}{n-k-2} \quad (2.24)$$

där  $\hat{\sigma}_k^2$  ges av (2.22),  $k$  är antalet parametrar i modellen och  $n$  är urvalets storlek. I denna uppsats kommer AICc och SIC att användas, för de båda gäller att ett lägre värde anses som bättre än ett högre (Shumway, R. & Stoffer, D. 2006).

---

3 Även kallat *Bayes Informationskriterium*, BIC



Utöver AICc och SIC kommer även MAPE eller *Mean Absolute Percent forecasting Error* att användas som utvärderingsmedel. Till skillnad från de två informationskriterierna anges MAPE i procent och bestäms genom (Montgomery, 2006);

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{(y_t - \hat{y}_t(t-1))}{y_t} \right| 100 \quad (2.25)$$

För att avgöra om tidsserierna beskrivs av de funna modellerna utförs test på residualerna. Dessa ska vara oberoende normalfördelade med väntevärde 0 och konstant varians,  $\sigma^2$ . En qq-plot visar om normalfördelning föreligger och givet att detta är uppfyllt visar Ljung-Box *goodness-of-fit-test* om residualerna är oberoende (Montgomery, 2006). Test är utformat så att det under  $H_0$  testar för om residualerna är fördelat oberoende. Test statistikan skriv som;

$$Q_{LB} = T(T+2) \sum_{k=1}^K \left( \frac{1}{T-k} \right) r_k^2 \quad (2.26)$$

Där  $T$  är antal observation i urvalet,  $r_k^2$  är autokorrelationer vid avstånd  $k$ ,  $Q_{LB} \sim \chi^2_{a, K-p}$  där  $p$  är antalet parametrar i modellen.

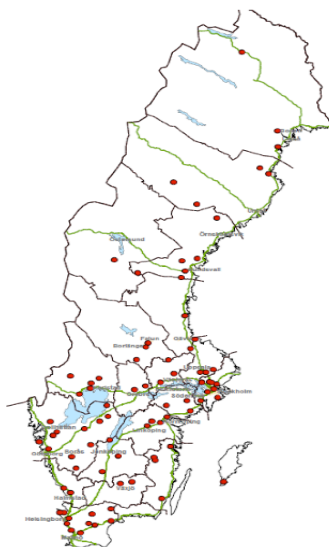
### 3 Datamaterial

#### 3.1 Trafikdata – procentuell förändring i mängden tung trafik

Datamaterialet har tillhandahållits av Trafikverket där det används för att uppskatta belastningen på det Svenska vägnätet. Syftet är att avgöra om det planerade underhållet är tillräckligt eller om andra åtgärder krävs för att hålla vägarna i fortsatt gott skick. Insamlingen har pågått sedan 1973 och således finns de årliga förändringarna skattade sedan 1974. För analysen i denna uppsats har data från 1992 till och med 2011 använts. Under denna period har två ramar använts. Den ena sträcker sig från januari 1992 och den andra från maj 1994. Den äldre ramen (9912) innehåller således fler observationer än den nyare (413). Denna skillnad blir dock mindre då observationspunkterna i ram 413 framförallt består av fast apparatur som ligger skyddad i vägbanan tillskillnad från observationspunkterna i ram 9912 som till största del består av temporära slangar. Som en konsekvens av detta är bortfallen i ram 413 betydligt färre och det slutgiltiga antalet observationer mer eller mindre detsamma. Observationerna används sedan för att skatta den procentuella förändringen i mängden trafik jämfört med samma månad föregående år.

I Figur 3.1 nedan kan punkternas ungefärliga placering i ram 413 ses.

Figur 3.1 Helårspunkter för skattning av trafikarbetets förändring<sup>4</sup>



##### 3.1.1 Stratifiering

Nuvarande ram består av c:a 22000st *trafikmätningssavsnitt*. Dessa avsnitt utgörs av oavbrutna sträckor väg, dvs sträckor mellan korsningar eller dylikt exklusive vägar med kommunalt underhåll som ligger innanför stadsplanegräns. Trafikmätningssavsnitten är uppdelade efter regioner (6) och inom regionerna är de uppdelade efter vägkategori (4). Detta ger totalt 24 st *huvudstratum*. Dessa strata är i sin tur uppdelade efter vilken trafiktyp som dominerar det aktuella avsnittet. Denna sista uppdelning resulterar i 83 st *finstrata*.

Ett trafikmätningssavsnitt per finstrata väljs sedan slumpmässigt ut för mätningen. Denna urvalsprocess är dock inte ett obundet slumpmässigt urval då punkterna viktas efter tidigare observerad trafikmängd. En punkt där mängden trafik tidigare har varit hög har större sannolikhet att väljas.

<sup>4</sup> Trafikarbetets förändring 2009-2010, Trafikverket

STOCKHOLMS UNIVERSITET  
Statistiska Institutionen

Vid varje mätpunkt observeras sex olika kategorier av trafikflöde uppdelade efter fordonstyp, personbil, lätt lastbil eller tung lastbil<sup>5</sup>. Dessa är sedan indelade i två underkategorier, med eller utan släp.

### 3.1.2 Skattningsmodell

För att skatta den procentuella förändringen i mängden trafik används en modell som dels skattar punktskattningar men även variansen för varje punktskattning.

$$\hat{R}_{TOT} = \frac{1}{\sum_{g=1}^G T_g} \sum_{g=1}^G \left( \frac{T_g}{n_g} \sum_{p=1}^{n_g} R_{gp} \right) \quad (3.1)$$

- där
- $n_g$  = Antal godkända punkter i variansgrupp  $g$
  - $R_{gp}$  = Förändring i flöde mellan  $t$  och  $t-1$  för punkt  $p$  i variansgrupp  $g$
  - $T_g$  = Vikt för både godkända och underkända punkter i variansgrupp  $g$
  - $G$  = Totala antal variansgrupper  $g$

En variansgrupp är en grupp av observationspunkter som uppvisar likartade egenskaper.

Variansen skattas enligt;

$$V[\hat{R}_{TOT}] = \sum_{g=1}^G \left( \frac{T_g}{\sum_{g=1}^G T_g} \right) \frac{1}{n_g(n_g-1)} \sum_{p=1}^{n_g} (R_{gp} - \bar{R}_g)^2 \quad (3.2)$$

där

$$\bar{R}_g = \frac{1}{n_g} \sum_{p=1}^{n_g} R_{gp} \quad (3.3)$$

Ett problem med modellen är att den bara använder observationer från direkt jämförbara timmar mellan observationer från  $t$  och  $t-1$ . D.v.s. endast tvärsnittet från de två perioderna används även om mer data finns tillgängligt.

I den mjukvara som användes för att genomföra skattningarna angavs som kriterium att minst 30% av de aktuella tidsperioderna skulle vara jämförbara.

---

<sup>5</sup> Definieras som lastbil med tre eller fler axelpar

STOCKHOLMS UNIVERSITET  
Statistiska Institutionen

### 3.1.3 Transformation

För att de två tidsserierna för trafik och BNP ska jämförbara har de skattade förändringarna i trafiken transformerats från månadsdata till kvartalsdata vilket är den finaste tillgängliga nivån över Sveriges BNP.

Den transformation som används är geometriska medelvärden då dessa är mer lämpade än t.ex. aritmetiska när genomsnittliga förändringar ska bestämmas.

Utöver kvartalsdata kommer även obehandlade observationer samt tre stegs glidande geometriska medelvärdesdata att analyseras, totalt analyseras alltså tre olika dataset för trafikförändringar innehållandes fyra variabler var.

Då datamaterialet är baserat på observationer i en förhållandevis besvärlig miljö saknas observationer för vissa kortare perioder. Med ett undantag är ingen av dessa perioder längre än en månad. Det lite längre gapet förekommer på ett sådant tidigt stadium att det inte heller skapar några större problem. Vid framställningen av de geometriskt glidande medelvärdes observationer för månader och kvartal har dessa saknade punkter uteslutits.

### 3.2 Nationalräkenskaper – procentuell förändring i BNP

Data över Sveriges BNP hämtades från Statistiska Centralbyrån (SCB) och omfattar den procentuella tillväxten i svenskt BNP från första kvartalet 1994 fram till andra kvartalet 2011. Detta datamaterial är även kalenderjusterat och säsongrensat. Kalenderjusteringen görs då kvartalen innehåller olika antal arbetsdagar och således är underlaget för BNP olika kvartal för kvartal. Det aktuella kvartalet jämförs med ett "normal kvartal" för att rensa bort eventuella effekter av ett större eller mindre antal arbetsdagar<sup>4</sup>. Detta blir av extra vikt då jämförelse görs med samma kvartal föregående år. Om ex. påsken då inföll i kvartal 1 istället för kvartal 2 och detta inte togs med i skattningen av förändringen skulle tillväxten i kvartal 1 framstå som större och tillväxten i kvartal 2 som mindre än den faktiska tillväxten i de båda kvartalen.

Säsongrensningen genomförs för att ta bort effekter orsakade av konsumtionsmönster. Exempelvis är 1:a kvartalet betydligt mycket svagare än det 4:e då konsumenterna återhämtar sig efter julhandeln. Långa perioder av semester som under sommarmånaderna kan också påverka nivån och tillväxten i BNP.

Värt att notera är också det faktum att de BNP-siffror som presenteras av SCB i första hand är preliminära. Inte förrän efter två års närmare undersökning kan mer precisa siffror presenteras. Detta medför att den data som finns tillgänglig för 2010 och 2011 inte är av samma kvalitet som från tidigare år.

---

4 Statistiska Centralbyrån, Avdelningen för Ekonomisk statistik, 2003. Säsongrensning av Nationalräkenskaperna – Översikt. Statistiks Centralbyrån, s. 28

## 4 Resultat

Innan analysen av de obehandlade förändringsprocenten genomfördes några mindre justeringar innan data behandlades av SAS. Under proceduren "proc arima" ignorerar SAS bortfall tills punkter med värden finns att analysera i den aktuella serien. Dessa punkter används sedan för prediktioner enligt den specificerade ARIMA modell som angivits fram tills det att punkter som saknar värde upptäcks. Efter att en sådan punkt har observerats av programvaran startar processen om och alla prediktioner från denna punkt baseras på värden som observerats efter det saknade<sup>6</sup>. För att undvika allt för stor förlust av data har därför saknade mätvärden estimeras med geometriska medelvärden baserade på de två föregående och de två efterföljande observationerna.

### 4.1 Trafik

Vid ARIMA-analysen av datan användes en etablerad arbetsgång. Först undersöktes tidsseriens stationaritet med hjälp av autokorrelationsfunktionen, ACF. Om ACF uppvisade tecken på en icke-stationär serie, se fig. 2.1, genomfördes differentiering och därefter undersöktes ACF på nytt. Av de data som inte bestod av geometriska glidande medelvärden differentierades alla variabler.

Då kravet på stationaritet uppfyllts studerades ACF och PACF för den nu bearbetade serien för att avgöra antalet AR och MA komponenter som skulle komma att ingå i ARIMA-modellen. Utifrån de förhållningsregler och riktlinjer som presenterades i kap. 2 bestämdes sedan vilka modeller som bäst beskrev datamaterialet.

Totalt användes tre olika dataset för analysen av förändring i trafikmängd. I tabellen nedan sammanfattas kortfattat vilka modeller som bedömdes beskriva data på bäst. Underlaget för bedömningarna tillsammans med en förteckning över dataseten presenteras i appendix A1.

**Tabell 4.1 Sammanfattning av ARIMA-modeller för förändring i trafikmängd**

Dataset	Variabel	Ram	ARIMA (p,d,q)(P,D,Q) <sup>s</sup>
1	Total tung lastbilstrafik	413	(1,1,1)(1,1,1) <sup>12</sup>
1	Tung lastbilstrafik med släp	413	(1,1,1)(1,1,1) <sup>12</sup>
1	Total tung lastbilstrafik	9912	(1,1,1)(1,0,1) <sup>12</sup>
1	Tung lastbilstrafik med släp	9912	(1,1,1)(1,0,1) <sup>12</sup>
2	Total tung lastbilstrafik	413	(1,0,4)
2	Tung lastbilstrafik med släp	413	(2,0,7)
2	Total tung lastbilstrafik	9912	Na <sup>7</sup>
2	Tung lastbilstrafik med släp	9912	Na <sup>8</sup>
3	Total tung lastbilstrafik	413	(1,0,1)(4,0,0) <sup>1</sup>
3	Tung lastbilstrafik med släp	413	(1,1,1)
3	Total tung lastbilstrafik	9912	(2,1,1)(5,0,3) <sup>3</sup>
3	Tung lastbilstrafik med släp	9912	(2,1,1)(5,0,3) <sup>3</sup>

<sup>6</sup> [http://support.sas.com/documentation/cdl/en/etsug/60372/HTML/default/viewer.htm#etsug\\_arima\\_sect041.htm](http://support.sas.com/documentation/cdl/en/etsug/60372/HTML/default/viewer.htm#etsug_arima_sect041.htm)  
den 21/12-2011

<sup>7</sup> Stora problem med stationaritet

<sup>8</sup> Stora problem med stationaritet

STOCKHOLMS UNIVERSITET  
Statistiska Institutionen

## 4.2 BNP

Vid analysen av BNP användes samma arbetsgång som vid analysen av trafik.

*Tabell 4.2 Resultat av ARIMA-analys av BNP*

Variabel	ARIMA (p,d,q)(P,D,Q)
Procentuell ökning från föregående kvartal	(0,1,2)(8,0,4) <sup>s</sup>

## 5 Slutsats

Som tydligt framgår av tabell 4.1 finns det tillräckligt med information i det studerade datamaterialet för att ska kunna ligga till grund för tidsseriemodeller av ARIMA-typ. Vad som dock syns tydligt i appendix A1 där korrelations och residualplottar redovisas är att en del av de presenterade modellerna är i relativt dåligt skick. Endast ett fåtal uppnår en tillräckligt hög kvalitet för att passera en hårdare granskning. En utav de främst bidragande orsakerna tros vara att analyserat data inte är direkt observationer utan skattningar som erhållits genom att observerat data har behandlats av en Trafikverket utvecklad modell.

Vad gäller samvarians med BNP verkar det finnas vissa tendenser till likartat beteende men inget konkret har påvisats.

## 6 Diskussion

I detta avsnitt analyseras de i kapitel fem presenterade slutsatserna närmare tillsammans med briser och begränsningar i studien.

### 6.1 Datamaterial

Det datamaterial som använts i studien är mycket omfattande men saknar ej brister, detta gäller speciellt det material som finns tillgängligt för analys av trafikförändringar. Som det beskrivs i kap. 3.1 är de observationspunkter som ligger till grund för förändringskattningarna ej slumpmässigt utvalda vilket är grundläggande antagande för analys av ett urval. Vad detta får för effekter för datamaterialet i slutändan är svårt att bedöma då det selektiva urvalet görs för att undvika att punkter med låg trafik får allt för stort inflytande på den slutgiltiga skattningen. Med säkerhet går det dock att säga att denna urvalsmetod medför att sannolikheten för överskattningar ökar dramatiskt.

En annan påverkande faktor är typ av datamaterial. Det material som använts och analyserats i studien är inte obehandlat utan skattningar som baserats på observationer. Den modell som använts för att producera dessa skattningar är konstruerad på ett sådant sätt att underskattningar undviks. Detta får följden att överskattningar dock blir mer förekommande. Beteendet kommer sig av att det primära syftet är att ge Trafikverket en bra indikation på i vilka områden slitaget på vägbanorna och tillhörande installationer är som störst och inte att ge så precisa skattningar som möjligt.

För att jämföra förändringar BNP och trafik var det nödvändigt att transformera trafikdata från månadsdata till kvartalsdata då BNP endast finns på kvartalsbasis. Detta innebar ytterligare ett steg bort från de oblandade observationerna och således ytterligare informationsförlust.





## 6.2 Studiens genomförande

Vad gäller ARIMA-analys är jag lång ifrån en expert. För att genomföra denna studie har jag därför fått lita mig mot litteratur samt den, dock mycket begränsade erfarenhet jag fått under mina studier. Detta gör att det presenterade resultatet inte bör granskas som fullständig sanning utan mer som en indikator på hur en sådan skulle kunna te sig.

## 6.3 Vidare analysmöjligheter

Det finns helt klart möjligheter att vidare analysera det datamaterial som använts för studien i denna uppsats. Dels genom att analysera de på skattningarna baserade observationsdatan vilket förmodligen innehåller betydligt mer information men å andra sidan är avsevärt mer svårtolkat än de mer städade månadsskattningarna.

Det är också fullt möjligt att använda mer avancerade verktyg eller mer komplexa versioner av ARIMA-analys som t.ex. VARMAX för att påvisa ett mer direkt samband mellan förändring i BNP och Trafik.



STOCKHOLMS UNIVERSITET  
Statistiska Institutionen

## 7 Referenser

Bernhardsson, J., 2002. *Tradingguiden*. Fischer & Co, s. 89

Barhoumi, K. mfl., 2008., *Occasional Paper no 84 - Short-term forecasting of GDP using large monthly datasets: a pseudo real-time forecast evaluation exercise*. ECB, s.5.

Gujarati, D., 2004. *Basic Econometrics* (4ed) McGraw-Hill, s.797.

J. D. Cryer & Kung-Sik Chan, 2008. *Time Series Analysis With Applications in R*. Springer, s.92.

Lequillier, F. & Blades, D., 2006. *Understanding National Accounts*. OECD, s.399.

Kirchgassner G., Wolters J., 2007. *Introduction to Modern Time Series Analysis*. Springer, s.165, 65.

Montgomery, C., Jennings, C. & Kulahci, M., 2006. *Introduction to Time Series Analysis and Forecasting*. Wiley, s.257, 249, 151, 32, 256, 59, 51, 57.

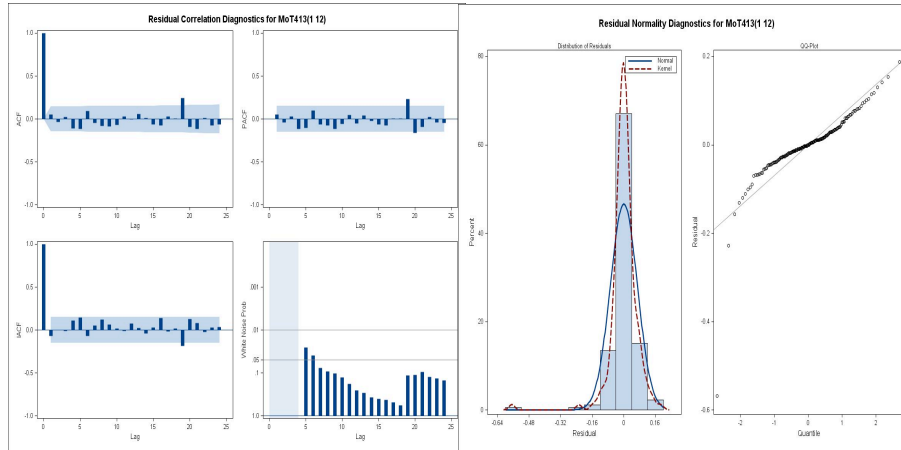
Shumway, R. & Stoffer, D., 2006. *Time Series Analysis and Its Applications* (2ed) Springer, s.23, 85, 86, 90, 93, 155, 156, 54, 53.

STOCKHOLMS UNIVERSITET  
Statistiska Institutionen

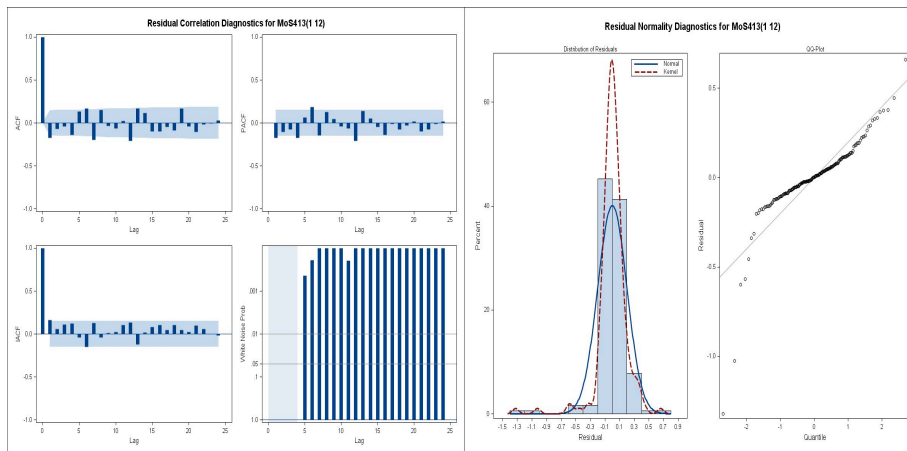
## Appendix 1 Korrelations- och residualplottar

1 – Månadsvisa observationer

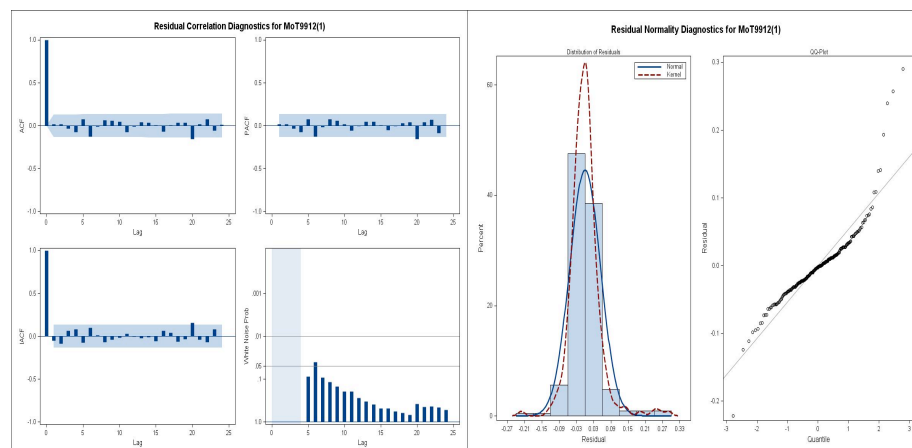
### Ram 413 – Total tung lastbilstrafik



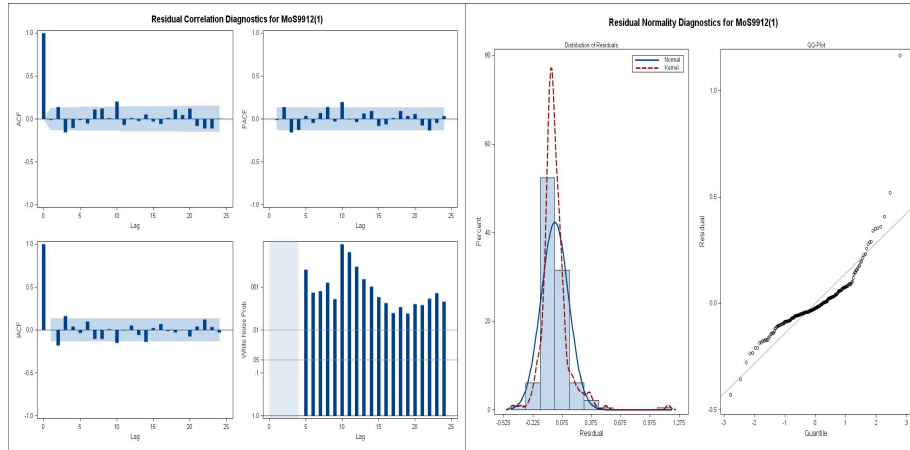
### Ram 413 - Tung lastbilstrafik med släp



### Ram 9912 – Total tung lastbilstrafik

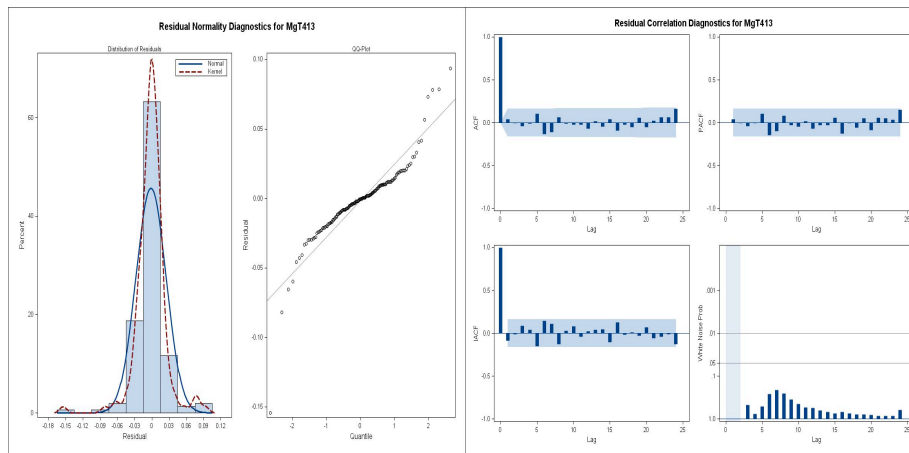


Ram 9912 - Tung lastbilstrafik med släp

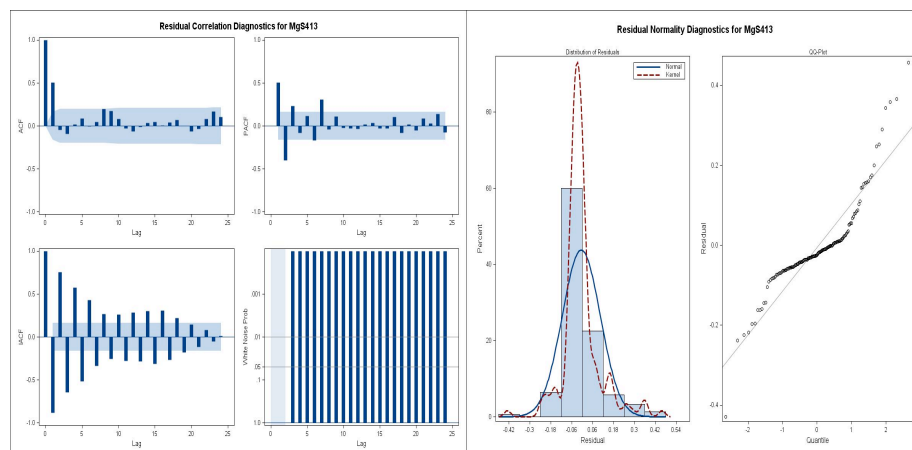


2 –Månadsvisa Glidande Geometriska Medelvärden

Ram 413 – Total tung lastbilstrafik



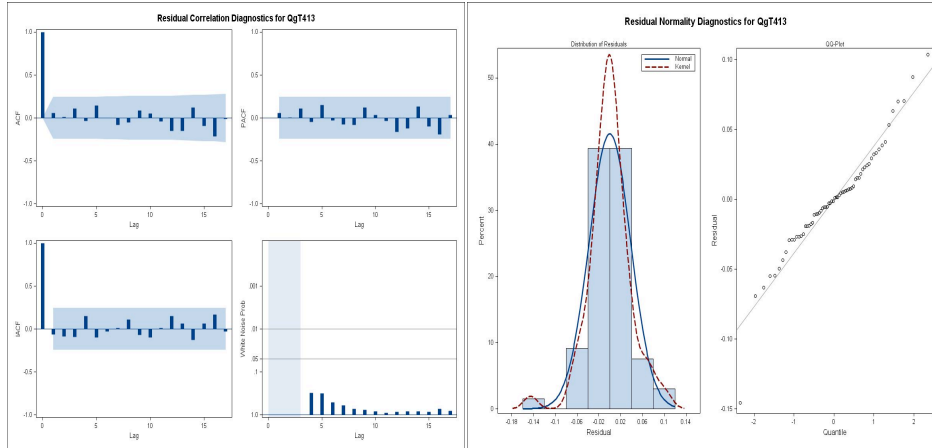
Ram 413 - Tung lastbilstrafik med släp



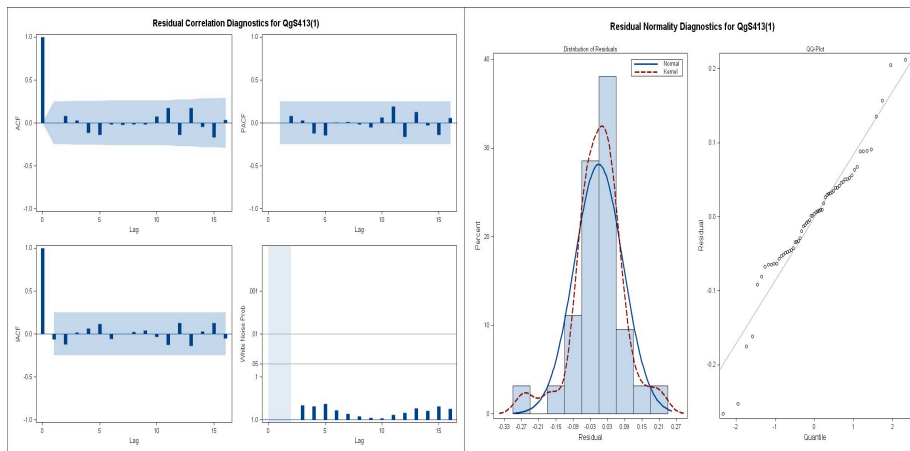
STOCKHOLMS UNIVERSITET  
Statistiska Institutionen

3 – Kvartalsvisa Geometrisk Medelvärden

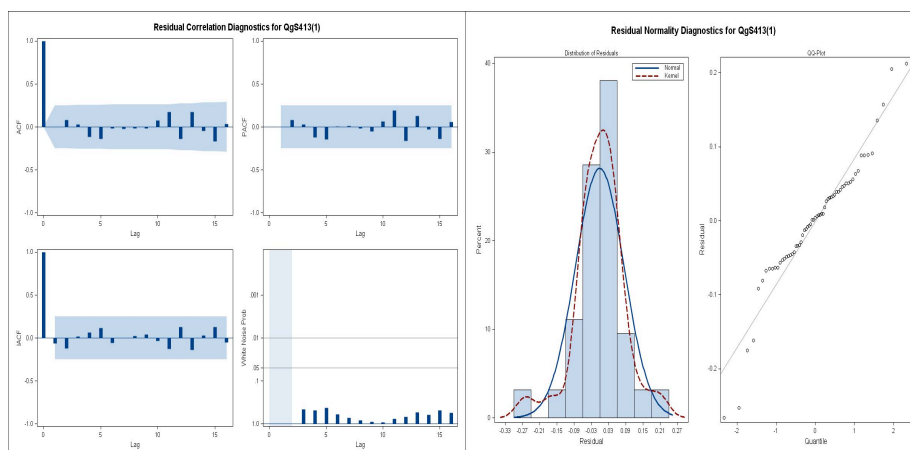
Ram 413 – Total tung lastbilstrafik



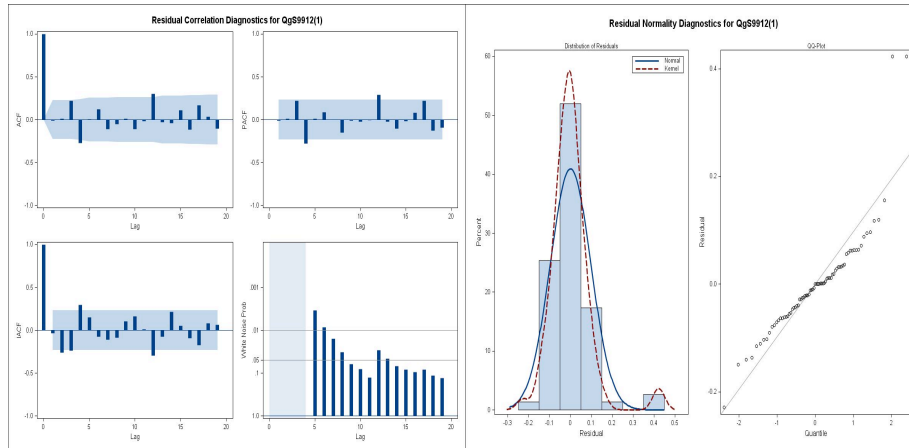
Ram 413 - Tung lastbilstrafik med släp



Ram 9912 – Total tung lastbilstrafik



Ram 9912 - Tung lastbilstrafik med släp



4 - BNP

