

2004:06

**Tidsserieanalys av svenska BNP-revideringar
1980–1999**

I serien Bakgrundsfakta presenteras bakgrundsmaterial till den statistik som avdelningen för ekonomisk statistik vid SCB producerar. Det kan röra sig om produktbeskrivningar, metodredovisningar samt olika sammanställningar av statistik som kan ge en överblick och underlätta användandet av statistiken.

Utgivna publikationer från 2001 i serien Bakgrundsfakta till Ekonomisk statistik

- 2001:1 Offentlig och privat verksamhet – statistik om anordnare av välfärdstjänster 1995, 1997 och 1999
- 2002:1 Forskar kvinnor mer än män? Resultat från en arbetstidsundersökning riktad till forskande och undervisande personal vid universitet och högskolor år 2000
- 2002:2 Forskning och utveckling (FoU) i företag med färre än 50 anställda år 2000
- 2002:3 Företagsenheten i den ekonomiska statistiken
- 2002:4 Statistik om privatiseringen av välfärdstjänster 1995–2001. En sammanställning från SCB:s statistikkällor
- 2003:1 Effekter av minskad detaljeringsgrad i varunomenklaturen i Intrastat – från KN8 till KN6
- 2003:2 Consequences of reduced grade in detail in the nomenclature in Intrastat – from CN8 to CN6
- 2003:3 SAMU. The system for co-ordination of frame populations and samples from the Business Register at Statistics Sweden
- 2003:4 Projekt med anknytning till projektet “Statistik om den nya ekonomin”. En kartläggning av utvecklingsprojekt och uppdrag
- 2003:5 Development of Alternative Methods to Produce Early Estimates of the Swedish Foreign Trade Statistics
- 2003:6 Övergång från SNI 92 till SNI 2002: Underlag för att bedöma effekter av tidsseriebrott
- 2003:7 Sveriges industriproduktionsindex 1913–2002 – Tidsserieanalys
The Swedish Industrial Production Index 1913–2002 – Time Series Analysis
- 2003:8 Cross-country comparison of prices for durable consumer goods: Pilot study – washing machines
- 2003:9 Monthly leading indicators using the leading information in the monthly Business Tendency Survey
- 2003:10 Privat drift av offentligt finansierade välfärdstjänster. En sammanställning av statistik
- 2003:11 Säsongrensning av Nationalräkenskaperna – Översikt
- 2003:12 En tillämpning av TRAMO/SEATS: Den svenska utrikeshandeln 1914–2003
- 2003:13 A note on improving imputations using time series forecasts
- 2003:14 Definitions of goods and services in external trade statistics

Fortsättning på omslagets tredje sida!

Ovannämnda rapporter, liksom övriga SCB-publikationer, kan beställas från:
Statistiska centralbyrån, SCB, Publikationstjänsten, 701 89 ÖREBRO,
telefon 019-17 68 00 eller fax 019-17 64 44.

Du kan också köpa SCB:s publikationer i **Statistikbutiken**:
Karlavägen 100, Stockholm

2004:06

**Tidsserieanalys av svenska BNP-revideringar
1980–1999**

Producent Statistiska centralbyrån
Avdelningen för Ekonomisk statistik

Förfrågningar Lars-Erik Öller, tel 08-506 943 33 (handledare)
Statistiska centralbyrån och Statistiska institutionen vid Stockholms
universitet
e-post: lars-erik.oller@scb.se

Uppsats på påbyggnadsnivå, höstterminen 2003
Jonas Nilsson
Olof Rosander

© 2004 Statistiska centralbyrån

ISSN 1650-9447

Printed in Sweden

SCB-Tryck, Örebro 2004.03  MILJÖMÄRKT Trycksak 341590

Innehåll

Förord	5
Sammanfattning	7
1 Inledning	9
1.1 <i>Bakgrund</i>	9
1.2 <i>Syfte</i>	9
1.3 <i>Data och metod</i>	9
1.4 <i>Disposition</i>	10
2 Teori	11
2.1 <i>Stationär tidsserie</i>	11
2.1.1 Dickey-Fuller test	12
2.1.2 Åtgärder mot icke-stationäritet	12
2.2 <i>Autokorrelation</i>	12
2.3 <i>Modellering av en tidsserie</i>	13
2.3.1 ARIMA-modell	14
2.3.2 Identifiering av en ARIMA-modell	15
2.4 <i>Diagnostik av modell</i>	16
2.4.1 Test av autokorrelation i residualerna	16
2.4.2 Test för icke-linjäritet	17
2.4.3 Normalitet.....	17
2.4.4 Kriterium för modellval vid flera tänkbara modeller	17
3 Svenska BNP-revideringar 1980–1999	19
4 Resultat och analys	22
4.1 <i>Autokorrelationsfunktioner</i>	22
4.2 <i>Modellskattningar</i>	24
4.3 <i>Vad betyder resultaten?</i>	24
5 Diskussion och slutsatser	26
Litteraturförteckning	27
APPENDIX 1 – autokorrelationsfunktioner för residualer	28

Förord

Denna uppsats kan betraktas som ett tillägg till Bilaga 4 i utredningen "Utveckling och förbättring av den ekonomiska statistiken", SOU 2002:118.

Uppgiften har varit att kontrollera för systematiska fel i BNP-revideringarna. Vi har delat upp arbetet genom att Olof har ägnat mer tid åt bakgrund och teori medan Jonas har koncentrerat sig mer på resultatdelen. Givetvis ansvarar vi för uppsatsen gemensamt samt eventuella felaktigheter i denna. Ett stort tack riktas till vår handledare Lars-Erik Öller vid SCB som givit värdefulla synpunkter och varit behjälplig med data.

Jonas Nilsson

Olof Rosander

Sammanfattning

Syftet med uppsatsen är att undersöka om tidsserierna för de svenska BNP-revideringarna 1980–1999 innehåller systematiska fel i form av autokorrelation samt att försöka identifiera och skatta modeller för dessa. Vi utgår från den metod för modellering av tidsserier som presenteras i Box & Jenkins (1976) och vi identifierar ARIMA-modeller där serierna korrigerats för extremvärden.

Resultatet visar att alla serier inte är autokorrelerade, däribland den totala BNP-serien. De serier som visar på autokorrelation är privat och offentlig konsumtion, investeringar, export, import och import av tjänster. Värt att påpeka är att exportrevideringarna visar på ett visst säsongsb beroende. De skattade modellerna kommer dock inte att kunna användas till att förbättra de preliminära siffrorna. Detta till följd av att den reviderade siffran för ett kvartal inte finns tillgänglig förrän två år efter att den preliminära siffran har presenterats. Vidare kan nämnas att hanteringen av extremvärden har stor påverkan på erhållna resultat.

1 Inledning

1.1 Bakgrund

I takt med att behovet och betydelsen av att kunna göra tillförlitliga prognoser av den framtida makroekonomiska utvecklingen har ökat, har även kraven på de data prognoserna måste bygga på ökat. Detta såväl med avseende på korrekthet som på hur snabbt publicerade data finns tillgängliga. För detta ändamål sammanställer SCB data från ett stort antal områden i den svenska ekonomin och i nationalräkenskaperna sammanställs och publiceras de delar som faller under det samlande begreppet bruttonationalprodukt (BNP).

Nationalräkenskaperna har av denna anledning blivit ett mycket användbart verktyg vid makroekonomiska analyser och prognoser. Detta ställer för SCB som statlig myndighet om möjligt än högre krav på att de siffror som publiceras är tillförlitliga. Även om det ideala vore att siffror publiceras snabbt med stor tillförlitlighet, är en avvägning mellan snabb tillgänglighet och tillförlitlighet i praktiken en ofrånkomlighet. Att samla in och sammanställa data från stora delar av ekonomin är i sig en svår uppgift. Om siffror skall kunna finnas tillgängliga för användare i ett tidigt skede, måste dessutom de preliminära siffrorna i många fall baseras på stickprovsundersökningar, interpoleringar eller extrapoleringar av tidigare data etc. Sett ur denna aspekt är de preliminära siffrorna i sig att betrakta som approximationer. Behovet av att revidera redan publicerade siffror när ytterligare data från ekonomin blir tillgängliga är därför ofta stort, vilket har medfört att tillförlitligheten i de preliminära siffrorna ibland har ifrågasatts. För en ingående diskussion av betydelsen av osäkerheten i preliminära siffror se Öller & Hansson (2002).

Är användarna rätt ute om de ifrågasätter tillförlitligheten i de preliminära siffrorna? Om de reviderade och de preliminära siffrorna ger en rätt olik bild av utvecklingen i ekonomin förfaller användare gör rätt i att använda de senare med viss försiktighet. Även om de preliminära siffrorna ofta är approximationer och bygger på en mindre mängd data än de reviderade siffrorna, finns det ingen anledning att förvänta sig att det resulterar i en systematisk bias i de reviderade siffrorna jämfört med de preliminära siffrorna.

Öller & Hansson (2002) undersöker om det förekommit en sådan bias mellan preliminära och reviderade svenska BNP-siffror mellan 1980–1998. Författarna finner att svenska revisioner, mätt som differensen mellan reviderad och preliminär siffra, ofta uppvisar en positiv bias dvs. de preliminära siffrorna ligger i genomsnitt lägre än de reviderade siffrorna. De finner vidare att revisionernas fördelningar i många fall är skeva och ofta har tjocka svansar. Problem med extremvärden föreligger även i flera av de undersökta variablerna.

Med utgångspunkt i ovan nämnda studie undersöker vi om svenska BNP-revideringar även uppvisar en annan form av systematiska fel; autokorrelation i tidsserierna för dessa variabler. Öller & Hansson (2002, s.21) påpekar att det bästa alternativet består i att förbättra de preliminära siffrorna genom att göra metoderna för statistikproduktionen bättre. Därför kan tecken på systematiska fel vara en indikation på att statistikproduktionen inte fungerar på ett optimalt (rationellt) sätt, se *ibid.* för en mer utförlig diskussion av vilka variabler som är i störst behov av en översyn och hur produktionen kan/bör förbättras.

1.2 Syfte

Syftet med denna uppsats består i att undersöka om det föreligger autokorrelation i revideringarna av de preliminära svenska BNP-siffrorna och om möjligt skapa modeller för de tids-serier där autokorrelation föreligger.

1.3 Data och metod

Vi använder i uppsatsen kvartalsdata från SCB över de svenska BNP-revideringarna 1980–1999. De preliminära respektive reviderade siffrorna är angivna som procentuell förändring från föregående kvartal i fasta priser. Varken de preliminära eller de reviderade siffrorna är säsong- eller arbetsdagkorrigerade.

En revidering är därmed definitionsmässigt skillnaden mellan en reviderad och preliminär tillväxtsiffra angiven i procentenheter. Beroende på att de preliminära siffrorna, som i regel publiceras 90 dagar efter kvartalsslut (SOU 2002:118, s.27), revideras fler än en gång måste man vid en undersökning av detta slag besluta sig för vilken av revideringarna som ska utgöra de "slutgiltiga" siffrorna. Det är dock viktigt att poängtera att de reviderade siffrorna i likhet med de preliminära siffrorna utgör approximationer. Detta eftersom man aldrig kan erhålla de exakta siffrorna för de variabler som ingår i BNP-måttet. Därför blir en studie av revideringar i stort en jämförelse mellan approximativa siffror, där vi dock förväntar oss att de reviderade siffrorna i regel ligger närmare de "sanna" siffrorna (Öller & Hansson [2002, s. 127]).

Vi väljer i denna uppsats att använda de reviderade kvartalssiffror som publiceras i slutet av november $t + 2$, dvs. två år efter det aktuella året, som våra slutgiltiga siffror. I likhet med Öller & Hansson (2002, s.123) gör vi detta dels för att senare revideringar ofta berör definieringar och metodval dels för att vi skall kunna använda siffror som reviderats på liknande sätt för hela den aktuella perioden.

Med utgångspunkt i tidsserier över revideringarna följer vi i stort den metod som presenteras i Box & Jenkins (1976, s.19). Denna metod bygger på att de observerade serierna är stationära (mer om stationaritet i *Avsnitt 2.1*) och för serier som uppfyller detta villkor beräknar vi autokorrelationsfunktioner. Utifrån dessa identifierar och estimerar vi ARIMA (Auto Regressive Integrated Moving Average)-modeller i de fall serierna visar på signifikant autokorrelation. Modellernas lämplighet testas sedan med olika diagnostiska test som kontrollerar att residualerna uppfyller modellantagandena. Ser modellernas residualer ut som vitt brus, återstår ingen dynamisk systematik i serierna. Visar diagnostiken däremot att en modell inte är adekvat modifieras modellen och en ny modell identifieras och estimeras.

Vi känner att det här även är nödvändigt med en mer utförlig kommentar av hur vi hanterar extremvärden i serierna. Detta eftersom vi inte kommer att använda oss av någon etablerad statistisk metod för att identifiera och korrigera extremvärden. Om serierna innehåller många värden som avviker kraftigt från de övriga observationernas mönster kommer modellskattningarna sannolikhet bli missvisande. Därför har vi valt att acceptera maximalt fem procent extremvärden i serierna, vilket i de aktuella serierna innebär två observationer.

Extremvärden identifieras genom att undersöka residualerna till de skattade modellerna och de observationer för vilka residualerna är betydligt större än två standardavvikelser korrigeras (Franses [1996, s.11]). Om serierna innehåller fler än två extremvärden korrigerar vi de som har störst residualer. Denna korrigering kan genomföras genom att inkludera dummyvariabler vid skattningen av modellen (Franses [1996, s.11]). Vår programvara tillåter oss dock inte att göra detta och därför korrigerar vi dessa observationer genom att sätta dem till medelvärdet för den observerade serien. Detta är ekvivalent med att säga att om vi inte visste vilka värden extremvärdena antar är vår bästa gissning att sätta dessa till det observerade medelvärdet. Nackdelen med detta förfarande är att vi jämnar ut serierna, vilket skulle kunna leda till att de justerade serierna uppvisar autokorrelation när de ursprungliga inte gör det. Modeller för de korrigerade serierna skattas sedan på nytt. Denna enkla metod att identifiera och korrigera extremvärden är något godtycklig. Önskvärt vore att vi kunde ha identifierat extremvärden i serierna med ett lämpligt program t.ex. TRAMO. Vi anser dock att även denna hantering av extremvärden kan vara motiverad, utan vilken eventuella resultat riskerar att i allt för stor utsträckning påverkas av extrema observationer.

1.4 Disposition

I nästföljande kapitel presenterar vi den statistiska teori vi använder i uppsatsen. I kapitel tre ger vi en kortfattad beskrivning av tidsserierna för BNP-variablerna samt undersöker om serierna uppfyller de nödvändiga villkoren för stationaritet. Därefter presenterar vi i kapitel fyra en analys av de serier som bedöms vara stationära. I kapitel fem följer en kort diskussion av slutsatser och resultat.

2 Teori

2.1 Stationär tidsserie

En grundläggande förutsättning för att kunna analysera en tidsserie genom ARIMA-modellering, är att den uppfyller kraven på svag stationaritet. Detta begrepps innebörd är att den underliggande sannolikhetsfördelningen inte ändrar sig med tiden. Definitionen av svag stationaritet är:

1. Medelvärde är konstant, dvs. om $E(Y_t) = \mu$.
2. Kovariansen mellan Y_t och Y_s endast beror av tidsdifferensen, dvs. om $C(Y_t, Y_s) = f(t - s)$.
3. Från förutsättning "2" följer att variansen existerar och är konstant, dvs. att $E(Y_t - \mu)^2 = \sigma^2$. Detta eftersom $\sigma^2 = V(Y_t) = C(Y_t, Y_t)$.

En vanlig metod för att undersöka om en observerad tidsserie är stationär består i att göra en visuell bedömning av den ursprungliga serien. En annan metod består i att studera seriens autokorrelationsfunktion. Om denna tenderar att dö ut långsamt är serien förmodligen inte stationär. De visuella metoderna kan kompletteras med ett parametriskt test där man kontrollerar om serien innehåller en s.k. enhetsrot. Om serien innehåller en enhetsrot är tidsserien inte stationär. (Gujarati [1999])

Vi tar som exempel serien

$$y_t = \phi y_{t-1} + a_t \quad (2.1)$$

där a är $N \sim iid(0, \sigma^2)$. Om vi antar att ϕ är ett och att serien vid tidpunkten noll antar värdet γ_0 , kan vi skriva tidsserien som

$$y_t = y_{t-1} + a_t = y_{t-2} + a_t + a_{t-1} = \dots = \gamma_0 + \sum_{i=1}^t a_i \quad (2.2)$$

Om vi analyserar *ekvation (2.2)* ser vi att det förväntade värdet av y_t är stationärt.

$$E(y_t) = E(\gamma_0 + \sum_{i=1}^t a_i) = E(\gamma_0) + 0 = \gamma_0$$

Men det är däremot inte variansen, eftersom

$$V(y_t) = V(\gamma_0 + \sum_{i=1}^t a_i) = V(\gamma_0) + V(\sum_{i=1}^t a_i) = 0 + tV(a) = t\sigma^2$$

Detta betyder att serien innehåller en enhetsrot. Man kan utöka denna modell till en slumpvandring med en driftparameter. Då blir varken det förväntade värdet av y_t eller dess varians stationära.

$$y_t = \mu + y_{t-1} + a_t = 2\mu + y_{t-2} + a_t + a_{t-1} = \dots = t\mu + \gamma_0 + \sum_{i=1}^t a_i \quad (2.3)$$

$$E(y_t) = E(t\mu) + E(\gamma_0) + E(\sum_{i=1}^t a_i) = t\mu + \gamma_0$$

$$V(y_t) = V(t\mu) + V(\gamma_0) + V(\sum_{i=1}^t a_i) = 0 + 0 + tV(a) = t\sigma^2$$

Anledningen till att det är viktigt att försäkra sig om att en tidsserie är stationär innan man börjar bygga en modell för prognoser, är att resultaten annars kan bli missvisande.

För att testa om serien innehåller en enhetsrot eller en driftparameter kommer vi att använda oss av ett Dickey-Fuller test. Testet går ut på att

1. avgöra om parametern ϕ i *ekvation (2.1)* är ett.
2. avgöra om driftparametern, μ , i *ekvation (2.3)* är skild från noll.

Normalt skulle man kunna använda ett vanligt t -test för detta, men eftersom t -testet endast visar riktiga resultat om serien är stationär, fungerar det inte i dessa situationer. Då kan man istället använda detta Dickey-Fuller test, vars värden tabulerats med hjälp av simulering.

2.1.1 Dickey-Fuller test

Först gör man om *ekvation (2.1)* till en differens (subtraherar med y_{t-1}).

$$\nabla y_t = y_t - y_{t-1} = \phi y_{t-1} - y_{t-1} + u_t = -(1 - \phi)y_{t-1} + a_t = \lambda y_{t-1} + a_t$$

Om ϕ är ett, kommer λ att vara noll. Om vi kan ha anledning att tro (genom testet) att denna parameter är skild från noll, är också ϕ skild från ett. Detta skulle då betyda att serien är stationär. Men vi kan också undersöka om serien innehåller en driftparameter. Då läggs interceptet till och modellen blir

$$\nabla y_t = \mu + \lambda y_{t-1} + a_t$$

I denna test kontrolleras dessutom om μ -parametern är skild från noll, givet att λ är noll (se Enders [1995, s.223]). En viktig förutsättning i detta test är att felen inte innehåller någon autokorrelation. Om så skulle vara fallet får man ta med fler tidsförskjutningar (*lags*) på differensen, ∇y_t . Då skulle modellen bli

$$\nabla y_t = \mu + \lambda y_{t-1} + \beta_1 \nabla y_{t-1} + \beta_2 \nabla y_{t-2} + \dots + \beta_p \nabla y_{t-p} + a_t$$

där p är antalet tidsförskjutningar av y som behövs för att ta bort eventuell autokorrelation i serien. Antal längsta tidsförskjutningar, p , bestäms genom att jämföra diagnostik för modeller med olika värden på p . De kritiska gränserna till detta test skiljer sig från de normala t -testerna. För en utförligare beskrivning samt tabeller över kritiska gränser se Enders (1995).

2.1.2 Åtgärder mot icke-stationäritet

Om en serie inte klarar förutsättningarna för stationäritet finns det åtgärder som rättar till ett sådant problem. Ett sätt är att bilda differenser, med t.ex. en periods tidsförskjutning. Det gör en serie med en linjär trend stationär. Om serien istället innehåller en kvadratisk trend får man bilda ytterligare en differens. En serie kan även vara icke-stationär på säsongsnivå (t.ex. fyra kvartal eller tolv månader), då måste man göra differenser på säsongsnivå. Om serien istället visar olik varians kan man transformera serien med hjälp av logaritmering. Det har egenskapen att förminska stora tal mer, relativt små tal, vilket får till följd att serien jämnas ut. I vårt fall är serierna redan procentuella förändringar så detta problem är inte relevant. För att kunna modellera serier med olika varians över tiden skulle vi behöva skatta s.k. ARCH (Auto Regressive Conditional Heteroskedasticity)-modeller (Fiorentini & Maravall [1996]).

2.2 Autokorrelation

Om en tidsserie innehåller systematiska element kommer seriens observerade värden att vara autokorrelerade sinsemellan. Ett sätt att undersöka om en tidsseries värden är korrelerade är att beräkna och visuellt undersöka seriens autokorrelationsfunktion. Denna beskriver om det finns ett samband mellan de värden tidsserien antar vid olika tidpunkter. Genom att beräkna korrelationskoefficienten mellan y_t och y_{t-k} erhålls autokorrelationen för en tidsförskjutning med k enheter ("lag k ") teoretiskt enligt följande formel:

$$\rho_k = \frac{\text{Cov}(Y_t, Y_{t-k})}{V(Y_t)}, \quad k = 1, 2, \dots \quad (2.4)$$

Givet en observerad tidsserie skattas *ekvation (2.4)* med den observerade seriens autokorrelation för lag k enligt:

$$\hat{\rho}_k = \frac{\sum_{t=1}^{n-k} (y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum_{t=1}^{n-k} (y_t - \bar{y})^2}, \quad k = 1, 2, \dots, n-k+1, \text{ om seriens första värde är } y.$$

där \bar{y} är seriens skattade medelvärde.

Vi anger här även hur den partiella autokorrelationsfunktionen beräknas teoretiskt samt hur denna skattas. Detta eftersom den partiella autokorrelationsfunktionen har en praktisk tillämpning vid identifiering av modeller för en observerad tidsserie (se vidare *Avsnitt 2.3* nedan). Teoretiskt beräknas den partiella autokorrelationsfunktionen, ρ_{kk} , med hjälp av ρ_k enligt:

$$\rho_{kk} = \begin{cases} \rho_1 & k = 1 \\ \frac{\rho_k - \sum_{j=1}^{k-1} \rho_{k-1,j} \rho_{k-j}}{1 - \sum_{j=1}^{k-1} \rho_{k-1,j} \rho_j} & k = 2, 3, \dots \end{cases} \quad (2.5)$$

där $\rho_{kj} = \rho_{k-1,j} - \rho_{kk} \rho_{k-1,k-j}$, $j = 1, 2, \dots, k-1$

Den skattade autokorrelationsfunktionen $\hat{\rho}_k$ används för att erhålla en skattning av den partiella autokorrelationsfunktionen där ρ_k i *ekvation (2.5)* ovan ersätts med ett estimat, $\hat{\rho}_k$. Tolkningen av den partiella autokorrelationsfunktionen är inte lika enkel som fallet är med autokorrelationsfunktionen. Denna funktion kan tänkas spegla hur autokorrelationsfunktionen uppför sig när effekten av mellanliggande observationer elimineras vid beräkningen av autokorrelationen mellan y_t och y_{t-k} . (Bowerman & O'Connell [1993])

2.3 Modellering av en tidsserie

Om villkoren för svag stationaritet bedöms vara uppfyllda kan autokorrelationsfunktionen, respektive den partiella autokorrelationsfunktionen utnyttjas för att undersöka om en observerad tidsserie kan beskrivas av en ARIMA-modell. Presentationen nedan av grundragen i ARIMA-modellering av en tidsserie utgår från Box & Jenkins (1976), som utvecklade metoder för att identifiera, estimeras och diagnostisera ARIMA-modeller för prognostisering.

Vi inleder med att presentera modeller för en tidsserie som ursprungligen är stationär enligt kriterierna i *Avsnitt 2.1* ovan och inte uppvisar ett säsongsmönster. Vi nämner därefter kortfattat hur dessa kan utökas till att även modellera säsong samt tidsserier som i utgångsläget är icke-stationära men där serierna innehåller delar som uppvisar stora likheter s.k. homogen icke-stationaritet. I det sistnämnda fallet kan serierna göras stationära genom differentiering, se *Avsnitt 2.1*.

2.3.1 ARIMA-modell

Den generella beteckningen ARIMA-modell innefattar för en ursprungligen stationär tidsserie egentligen tre olika typer av linjära tidsseriemodeller: AR(p)-modeller, MA(q)-modeller samt ARMA(p, q)-modeller som är en kombination av de båda första. En autoregressiv modell av ordningen p , där beteckningen p anger hur många tidsförskjutningar av föregående värden som skall ingå i modellen, beskrivs med följande formel:

$$y_t = \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + a_t. \quad (2.6)$$

För att slippa skriva ut *ekvation (2.6)* förkortas denna vanligen som

$$\phi_p(B)y_t = a_t,$$

där $\phi_p(B) = 1 - \phi_1 B - \dots - \phi_p B^p$. B definieras som $B^k y_t = y_{t-k}$ för $k = 0, 1, 2, \dots$ och indikerar vilka av tidsseriens föregående värden som skall ingå i modellen.

Om en tidsserie kan beskrivas av en AR(p)-modell i enlighet med *ekvation (2.6)* ovan kommer det observerade värdet y_t att bero av tidsseriens p föregående värden, vart och ett multiplicerat med en koefficient specifik för respektive tidsförskjutning plus en slumpterm som fångar upp effekten av icke-observerade faktorer, vanligen benämnt vitt brus.

En glidande medelvärdesmodell av ordningen q , som anger vilka avvikelser från föregående perioder som ska ingå i modellen, beskrivs med följande formel:

$$y_t = a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q} = \theta_q(B)a_t \quad (2.7)$$

där $\theta_q(B) = 1 - \theta_1 B - \dots - \theta_q B^q$

Om en tidsserie kan beskrivas med en MA(q)-modell kommer det observerade värdet y_t att bero av föregående q slumpmässiga avvikelser multiplicerade med var sin koefficient.

Slutligen kan en modell bestå både av AR(p)- och MA(q)-komponenter som tillsammans bildar en ARMA(p, q)-modell (AutoRegressive Moving Average Model). En ARMA(p, q)-modell erhålls genom att kombinera *ekvation (2.6)* och (2.7) ovan enligt:

$$y_t = \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q} \quad (2.8)$$

alternativt

$$\phi_p(B)y_t = \theta_q(B)a_t$$

där $\phi_p(B) = 1 - \phi_1 B - \dots - \phi_p B^p$, $\theta_q(B) = 1 - \theta_1 B - \dots - \theta_q B^q$.

Dessutom krävs att modellens parametrar uppfyller stationaritets- och inverteringsvillkor (se Box & Jenkins [1976]). Om den ursprungliga tidsserien har ett medelvärde som är skilt från noll kan även en konstant inkluderas i *ekvationerna (2.6)*, (2.7) eller (2.8) ovan.

De ovan presenterade modellerna kan utökas till att även inkludera säsongsmönster i en tidsserie. Om en tidsserie uppvisar ett säsongsmönster kommer det att finnas ett beroende mellan seriens observerade värden på säsongsnivå (kvartal, månad) dvs. autokorrelation mellan y_t och y_{t-s} , där $s = 4$ för kvartals- och $s = 12$ för månadsdata. En ARMA(P, Q)-modell, där P och Q motsvarar p och q , på säsongsnivå, beskrivs med följande modell:

$$\Phi_P(B^s)y_t = \Theta_Q(B^s)\alpha_t, \quad (2.9)$$

där B definieras som $B^s y_t = y_{t-s}$.

ARMA(p,q)-modellen i *ekvation (2.8)* ovan ger tillsammans med säsongmodellen i *Ekvation 2.9* en ARMA(p,q)(P,Q)-modell enligt följande formel:

$$\phi_p(B)\Phi_P(B^s)y_t = \theta_q(B)\Theta_Q(B^s)a_t. \quad (2.10)$$

Om den ursprungliga serien är stationär är modellen i *ekvation (2.10)* tillräcklig. Är serien inte stationär inkluderas antalet nödvändiga differentieringar för att göra serien stationär, d på icke-säsongnivå och D på säsongnivå, och den generella ARIMA(p,d,q)(P,D,Q)-modellen får följande utseende:

$$\phi_p(B)\Phi_P(B^s)\nabla^d\nabla_s^D y_t = \theta_q(B)\Theta_Q(B^s)a_t,$$

där $\nabla^d = (1-B)^d$, $\nabla^D = (1-B^s)^D$. Parametrarna måste i likhet med övriga modeller uppfylla stationaritets- och inverteringsvillkoren.

2.3.2 Identifiering av en ARIMA-modell

För att avgöra om en tidsserie kan förklaras av en AR- och/eller MA-modell består ett möjligt tillvägagångssätt i att visuellt undersöka $\hat{\rho}_k$ och $\hat{\rho}_{kk}$ och jämföra dessa med ρ_k respektive ρ_{kk} för vissa teoretiska modeller. Vi ger här exemplet med en teoretisk AR(1)- och MA(1)-modell. AR(1)-modellen ges av:

$$y_t = \phi_1 y_{t-1} + a_t. \quad (2.11)$$

Genom att utveckla kovariansuttrycket i täljaren av *ekvation (2.4)* ovan får autokorrelationsfunktionen följande utseende för denna modell

$$\rho_k = \frac{\phi_1^k V(y_t)}{V(y_t)} = \phi_1^k. \quad (2.12)$$

Ekvation (2.12) visar att ρ_k för en AR(1)-modell avtar exponentiellt och går mot noll när k växer. AR(1)-modellen måste alltså uppfylla stationaritetsvillkoret att $|\phi_1| < 1$. Om $|\phi_1| > 1$ kommer $\rho_k \rightarrow \infty$, när $k \rightarrow \infty$.

En MA(1)-modell beskrivs med följande ekvation:

$$y_t = a_t - \theta_1 a_{t-1} \quad (2.13)$$

Vidare kan man med enkel algebra utveckla *ekvation (2.4)* ovan och visa att autokorrelationsfunktionen får följande utseende:

$$\begin{aligned} \rho_0 &= 1 \\ \rho_1 &= \frac{-\theta_1}{1+\theta_1^2} \\ \rho_k &= 0 \quad k = 2,3,\dots \end{aligned}$$

Detta visar att för en MA(1)-modell kommer autokorrelationsfunktionen sluta tvärt efter $k=1$. I MA(1)-modellen måste inverteringsvillkoret $|\theta_1| < 1$ vara uppfyllt.

Det är betydligt svårare att visa hur den partiella autokorrelationsfunktionen ser ut för de teoretiska AR(1)- respektive MA(1)-modellerna. Egenskaperna kommer dock vara omvända, vilket innebär att den partiella autokorrelationsfunktionen slutar tvärt vid $k=1$ för AR(1)-modellen och avtar exponentiellt för MA(1)-modellen, alltså tvärtom mot autokorrelationsfunktionen.

Vid identifiering av en ARIMA-modell uppvisar autokorrelationsfunktionen samt den partiella autokorrelationsfunktionen samma mönster på säsongsnivå som på icke-säsongsnivå. ARIMA-modeller med AR-komponenter (MA-komponenter) på både säsongs- och icke-säsongsnivå kommer att få en multiplikativ form. Vi exemplifierar med en ARIMA(1,0,0)(1,0,0)-modell.

$$(1 - \phi_1 B)(1 - \Phi_1 B^4) y_t = y_t - \phi_1 y_{t-1} - \Phi_1 y_{t-4} + \phi_1 \Phi_1 y_{t-5} = a_t \quad (2.14)$$

Ekvation (2.14) visar att modellen även inkluderar effekten av värdet för fem perioders tidsförskjutning. För en mer ingående beskrivning av ARIMA-modeller med säsong se t.ex. Box & Jenkins (1976).

2.4 Diagnostik av modell

För att säkerställa att en skattad ARIMA-modell beskriver en tidsserie på ett adekvat sätt kan modellen testas visuellt eller med hjälp av mer avancerade diagnostiska test som utvecklats i samma syfte. Såväl de visuella metoderna som de diagnostiska testarna utnyttjar modellens residualer för detta ändamål. Utöver några av dessa metoder/test presenterar vi två kriterier för val av modell när flera än en modell är möjlig.

2.4.1 Test av autokorrelation i residualerna

Autokorrelation i modellens residualer tyder på att den valda modellen inte beskriver tidsserien på ett riktigt sätt. Visuellt kan autokorrelation i residualerna undersökas genom att granska autokorrelationsfunktionen, respektive den partiella autokorrelationsfunktionen för residualerna. Om dessa visar att residualerna är korrelerade tyder detta på att modellen måste omspecificeras. Ett Ljung-Box test kan besvara denna fråga.

Ljung-Box testet visar om de m första autokorrelationerna för modellens residualer sammantaget ger stöd för att den valda modellen är rimlig. Teststatistikan ges av:

$$Q_m^* = n(n+2) \sum_{k=1}^m (n-k)^{-1} \hat{\rho}_k^2(\hat{a}) \quad (2.15)$$

där n är antalet residualer och $\hat{\rho}_k^2(\hat{a})$ är den skattade autokorrelationen för residualerna vid tidsförskjutning k . (Bowerman & O'Connell [1993])

Under nollhypotesen att residualerna inte är autokorrelerade kommer Q_m^* att följa en asymptotisk $\chi^2(m-b)$ -fördelning, där b är antalet skattade parametrar. Om Q_m^* är större än $\chi_\alpha^2(m-b)$ förkastas nollhypotesen på signifikansnivån α och modellen bör modifieras. Detta är ekvivalent med att förkasta nollhypotesen om det observerade p -värdet är mindre än α .

Franses (1996, s.9–10) påpekar att Ljung-Box testet gör det möjligt att identifiera en modell som är otillräcklig men saknar möjlighet att identifiera hur modellen bör modifieras. Skattas ARIMA-modeller utan tillgång till ett program som väljer den mest lämpliga modellen, bör därmed detta test kombineras med en granskning av autokorrelationsfunktionen (och eventuellt den partiella autokorrelationsfunktionen) för residualerna. Vidare riskerar autokorrelationer för korta tidsförskjutningar att inte få genomslag i testet om m i *ekvation (2.15)* ovan väljs för stort.

2.4.2 Test för icke-linjäritet

Även om residualerna inte visar på någon autokorrelation kan det finnas ett beroende mellan dem, vilket skulle kunna betyda att serien är icke-linjär (vilket även kan innebära heteroskedasticitet). En förutsättning för ARIMA-modellering är att serien är linjär samt att variansen är konstant över tiden. En metod att testa för detta är att undersöka om de kvadrerade residualerna är autokorrelerade. En möjlighet är att undersöka autokorrelationsfunktionen för de kvadrerade residualerna (Maravall [1983]). En annan är att beräkna Ljung-Box Q_m^{*2} , där $\hat{\rho}_k^2(\hat{a}^2)$ i ekvation (2.15) ersätts med:

$$\hat{\rho}_k^2(\hat{a}^2) = \left[\frac{\sum_{t=1}^{n-k} \left(\hat{a}_t^2 - \sum_{t=1}^n \hat{a}_t^2 / n \right) \left(\hat{a}_{t-k}^2 - \sum_{t=1}^n \hat{a}_t^2 / n \right)}{\sum_{t=1}^{n-k} \left(\hat{a}_t^2 - \sum_{t=1}^n \hat{a}_t^2 / n \right)^2} \right]^2$$

Ljung-Box Q_m^{*2} följer en asymptotisk $\chi_a^2(m-b)$ -fördelning och nollhypotesen, att de kvadrerade residualerna är okorrelerade, förkastas enligt samma förfarande som beskrivs i *Avsnitt 2.4.1* ovan.

2.4.3 Normalitet

Ett sätt att undersöka om modellens residualer är approximativt normalfördelade är att visuellt jämföra deras fördelning med normalfördelningen. Detta bygger på att sannolikheter för den teoretiska normalfördelningens observationer jämförs med sannolikheter för att modellens residualer kommer från en normalfördelning. Ett test som mäter hur långt ifrån det teoretiska normalfördelningsvärdet (ordinatan) residualerna ligger är Anderson-Darlings test-statistiska A2. Testet kvadrerar avståndet och ger större vikter till fördelningens svansar. Under nollhypotesen att residualerna är approximativt normalfördelade förkastas normalfördelningsantagandet för stora värden på A2.

Ett parametriskt test, främst för stora stickprov, är Jarque-Bera test (*JB*). Detta testar för avvikelser från normalfördelning i form av skevhet (*S*) och toppighet (*T*). Statistikan är asymptotisk χ^2 -fördelad med 2 frihetsgrader och beräknas enligt formeln:

$$JB = \frac{n}{6} \left[S^2 + \frac{(T-3)^2}{4} \right]$$

För beräkningsförfarande för skevhet och toppighet se Gujarati (1999, s.177–178). Under nollhypotesen att residualerna är approximativt normalfördelade förkastas normalfördelningsantagandet för stora värden på *JB*.

2.4.4 Kriterium för modellval vid flera tänkbara modeller

Om vi i utgångsläget har ett flertal möjliga modeller med olika antal skattade parametrar bör vi välja en modell som ger bäst anpassning med minsta antal skattade parametrar. Två statistikor som ofta används för att avgöra vilken av flera lämpliga modeller som bör väljas är Akaikes informationskriterium (AIC) och Schwarz-Bayes informationskriterium (BIC). Dessa kan approximeras enligt följande:

$$AIC(k) = n \log \hat{\sigma}_{ML}^2 + 2k$$

$$BIC(k) = n \log \hat{\sigma}_{ML}^2 + k \log n$$

där n är antal observationer, $\hat{\sigma}_{ML}^2 = SSE / n$ och k är antal skattade parametrar i modellen (Franses [1996 s.13]).

Den modell som har lägst värde på AIC och/eller BIC bör väljas utifrån dessa kriterier. Statistiska är främst utformade för modeller med många skattade parametrar, eftersom den straffar modeller hårdare ju fler parametrar som ingår. Om två lämpliga modeller i det senare fallet har samma antal skattade parametrar kommer vi att utnyttja modellens medelkvadratfel (MSE) som kriterium vid modellval. När modellerna har samma antal skattade parametrar och likvärdiga medelkvadratfel kan parametrarnas signifikans samt övriga diagnostiska test ge en bättre indikation på vilken modell som bör väljas.

3 Svenska BNP-revideringar 1980–1999

I detta kapitel ger vi först en övergripande bild av hur revideringarna av svenska BNP-siffror utvecklats sedan 1980-talet. Vi undersöker även om tidsserierna är stationära och därmed uppfyller villkoren i *Avsnitt 2.1*.

Diagram 3.1–3.13 nedan visar kvartalsdata för svenska BNP-revideringar 1980–1999. Startpunkterna för de olika serierna skiljer sig dock något åt mellan de olika variablerna. Detta till följd av att sammanhängande data för vissa variabler inte finns tillgängligt för hela perioden. Den ljusa linjen markerar medelvärdet.

Diagrammen visar tydligt att revideringarna för vissa variabler är mycket stora. (Läsaren bör ha i åtanke att revideringarna definieras som skillnaden mellan slutgiltig och preliminär tillväxtsiffra angett i procentenheter). Även revideringarna för total BNP är av betydande storlek. Samtliga observerade serier, fränsett lager, har ett medelvärde som skiljer sig från noll, även om detta inte testats för signifikans. Av dessa har alla variabler utom investeringar ett medelvärde större än noll. Denna bias i serierna undersöks ingående i den tidigare nämnda studien av Öller & Hansson (2002) och läsaren hänvisas till denna för en mer ingående analys för respektive variabel.

Diagram 3.9 och *3.12* visar att revideringarna för export respektive import av tjänster följer ett likartat mönster. I båda fallen har revideringarna minskat och fluktuerar inte lika kraftigt under 1990-talet som under 1980-talet, vilket också Öller & Hansson (2002) finner i sin studie. Det omvända gäller för statlig- respektive kommunal konsumtion i *Diagram 3.3* och *3.4*. För dessa fyra variabler bedömer vi med anledning av detta att de inte uppfyller stationaritetsvillkoret om konstant varians över tiden. Därför kommer ingen ytterligare analys av revideringarna för dessa variabler genomföras i nästföljande kapitel.

De flesta serier förefaller visuellt uppfylla det första stationaritetsvillkoret om konstant väntevärde över tiden. Trots att serierna är relativt korta valde vi att genomföra ett utökat Dickey-Fuller test på serierna för kommunal konsumtion (*Diagram 3.4*) och investeringar (*Diagram 3.5*), som visuellt möjligen inte uppfyller detta villkor. Resultatet visade att serierna inte innehåller enhetsrötter.

Om de största revideringarna dessutom utgör extremvärden i serierna måste dessa korrigeras för att inte påverka resultaten vid autokorrelations- respektive modellskattningarna. Den mycket stora negativa revideringen (–18,6 procent) för investeringar kvartal två 1995 i *Diagram 3.5* utgör ett tydligt extremvärde då serien i övrigt är relativt jämn. Beroende på att siffrorna för investeringar gällande kvartal två måste baseras på företagens investeringsplaner (vid kvartal ett för nästkommande kvartal) och inte på en stickprovsundersökning, är siffrorna för detta kvartal mycket osäkra. Av den anledningen väljer vi att korrigera värdet för revideringen kvartal två 1995 genom att sätta detta till seriens medelvärde redan innan vi skattar autokorrelationsfunktionen för serien.

Diagrammen 3.7 och *3.10* visar att både export och import har en revidering för kvartal två 1987 som är stor i förhållande till övriga värden i serierna. Detta hänger samman med att revideringarna för export respektive import av tjänster är mycket stora vid samma tidpunkt (se *Diagram 3.9* respektive *3.12*), vilket får genomslag i serierna för export respektive import. Revideringarna för kvartal två 1987 för export och import utgör därmed troligen extremvärden, se *Tabell 4.1* för hur värdena har korrigerats.

Diagram 3.1: Privat konsumtion
Revideringar kv.1 1980 - kv.4 1999

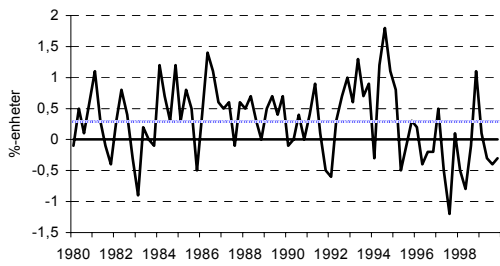


Diagram 3.2: Offentlig konsumtion
Revideringar kv.1 1980 - kv.4 1999

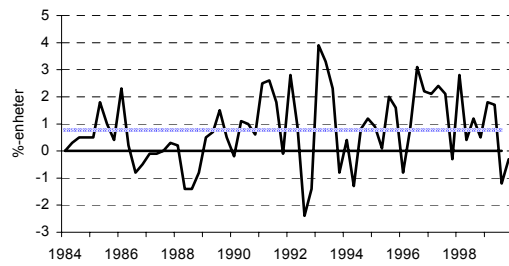


Diagram 3.3: Statlig konsumtion
Revideringar kv.2 1984 - kv.4 1999

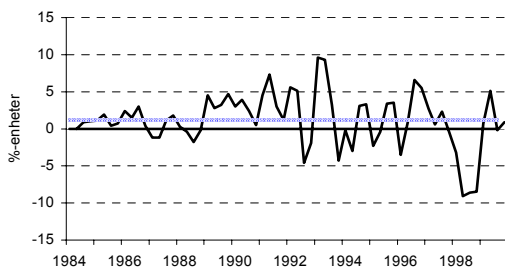


Diagram 3.4: Kommunal konsumtion
Revideringar kv.2 1984 - kv.4 1999

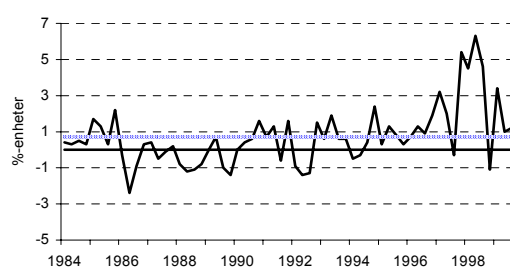


Diagram 3.5: Investeringar
Revideringar kv.2 1984 - kv.4 1999

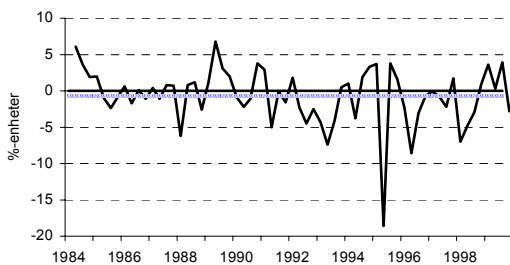


Diagram 3.6: Lager
Revideringar kv.2 1984 - kv.4 1999

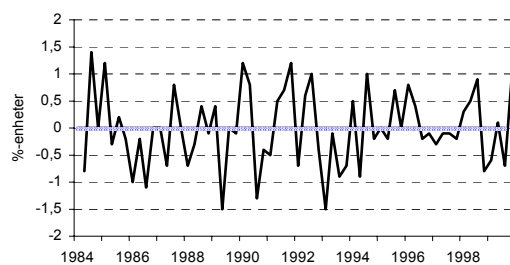


Diagram 3.7: Export
Revideringar kv.1 1984 - kv.4 1999

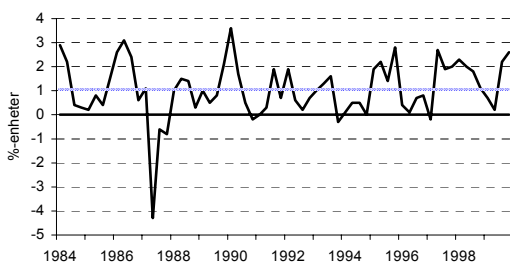


Diagram 3.8: Export av varor
Revideringar kv.1 1984 - kv.4 1999

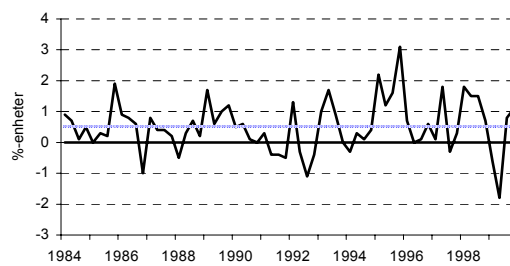


Diagram 3.9: Export av tjänster
Revideringar kv.1 1984 - kv.4 1999

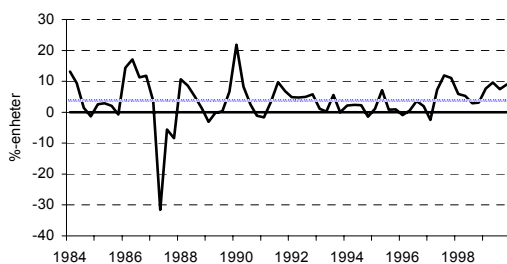


Diagram 3.10: Import
Revideringar kv.1 1984 - kv.4 1999

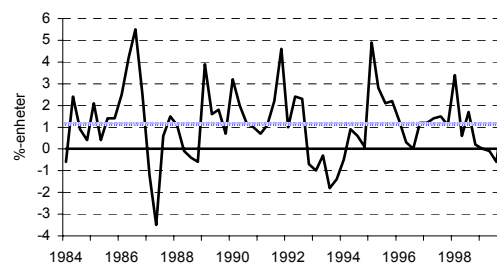


Diagram 3.11: Import av varor
Revideringar kv.1 1984 - kv.4 1999

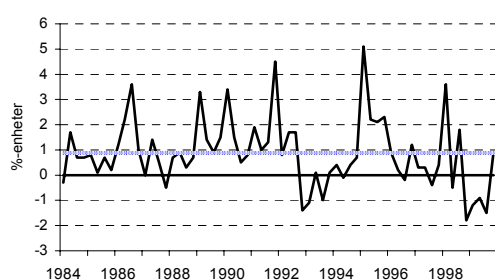


Diagram 3.12: Import av tjänster
Revideringar kv.1 1984 - kv.4 1999

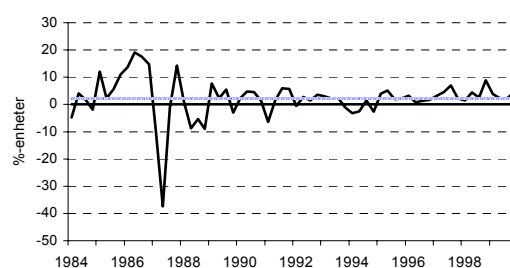
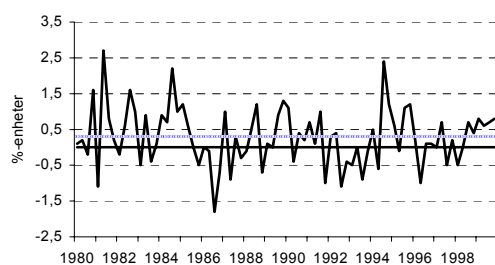


Diagram 3.13: BNP
Revideringar kv.1 1980 - kv.4 1999



4 Resultat och analys

4.1 Autokorrelationsfunktioner

De serier som uppfyller förutsättningarna för stationaritet kommer i detta avsnitt att testas för förekomst av autokorrelation. Tidsserier som inte innehåller autokorrelation är:

- BNP
- Lager
- Export av varor

I *diagram 4.1 – 4.6* nedan presenteras autokorrelationsfunktioner för privat och offentlig konsumtion, investeringar, export, import och import av varor. Dessa tidsserier visar på signifikant autokorrelation efter justering av extremvärden. Extremvärden för serierna har identifierats och justerats i enlighet med vad som redogörs för i *Avsnitt 1.3*. I samtliga serier frånsett privat konsumtion har extremvärden korrigerats och de korrigerade observationerna presenteras i *Tabell 4.1*.

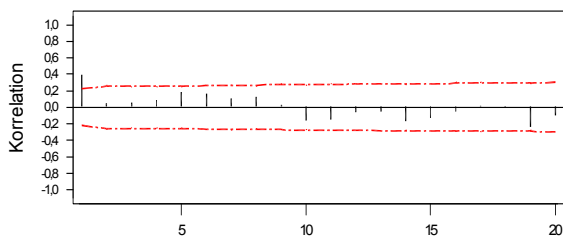
Tabell 4.1
Justerade extremvärden

Variabel	Observation	Revidering ^a	Residual ^a	s	Korrigerat värde ^a
Offentlig konsumtion	kv.3 1992	-2,4	-2,916	1,261	0,765
	kv.1 1993	3,9	3,538		0,765
Investering	kv.2 1995	-18,6	*	*	-0,710
Export	kv.2 1987	-4,3	-4,920	1,085	1,056
Import	kv.2 1987	-3,5	-3,692	1,5	1,142
	kv.1 1995	4,9	4,184		1,142
Import av varor	kv.4 1991	4,5	3,515	1,327	0,870
	kv.1 1995	5,1	4,282		0,870

a: Anges i procentenheter

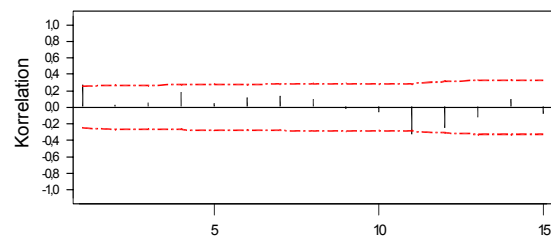
Diagram 4.1 – 4.6 visar på signifikant autokorrelation i serierna om en stapel passerar den streckade linjen i diagrammen. Det innebär att den skattade korrelationen, med 95 procents sannolikhet är skild från noll, om stationaritetvillkoren är uppfyllda. Diagrammen visar således att samtliga serier är autokorrelerade med ett kvartals tidsförskjutning. I diagrammen presenteras också (för varje kvartals tidsförskjutning) korrelationens skattade värde (*Korr*), dess *t*-värde samt ett värde för Ljung-Box Q_m^* (se *Avsnitt 2.3.1*). Diagrammen nedan utnyttas för att identifiera vilken typ av ARIMA-modell som är lämplig att skatta för respektive serie.

Diagram 4.1: Autokorelationsfunktion för privat konsumtion



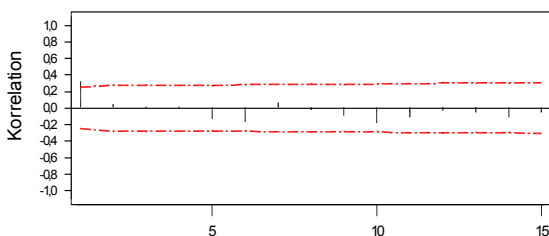
Lag	Korr	T	LBO	Lag	Korr	T	LBO	Lag	Korr	T	LBO
1	0,39	3,47	12,46	8	0,12	0,91	21,09	15	-0,13	-0,92	31,42
2	0,05	0,38	12,66	9	0,02	0,15	21,13	16	-0,05	-0,38	31,73
3	0,06	0,45	12,94	10	-0,17	-1,22	23,76	17	0,02	0,10	31,76
4	0,08	0,63	13,49	11	-0,15	-1,11	23,04	18	-0,00	-0,02	31,76
5	0,18	1,40	16,34	12	-0,06	-0,44	23,40	19	-0,24	-1,63	37,88
6	0,17	1,26	18,78	13	-0,06	-0,41	23,73	20	-0,10	-0,65	38,94
7	0,10	0,76	19,72	14	-0,17	-1,21	23,65				

Diagram 4.2: Autokorelationsfunktion för offentlig konsumtion



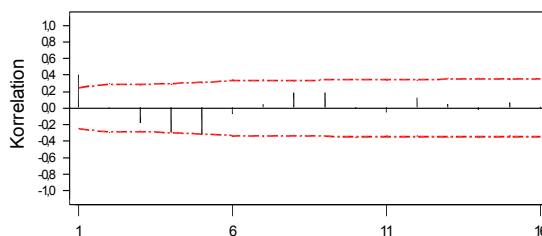
Lag	Korr	T	LBO	Lag	Korr	T	LBO	Lag	Korr	T	LBO
1	0,27	2,16	4,89	8	0,10	0,69	10,46	15	-0,09	-0,52	27,00
2	0,03	0,19	4,93	9	-0,02	-0,15	10,50				
3	0,06	0,43	5,16	10	-0,06	-0,43	10,79				
4	0,18	1,33	7,42	11	-0,33	-2,26	19,20				
5	0,04	0,29	7,54	12	-0,25	-1,62	24,33				
6	0,11	0,78	8,39	13	-0,12	-0,77	25,60				
7	0,14	0,97	9,74	14	0,10	0,59	26,37				

Diagram 4.3: Autokorelationsfunktion för investeringar



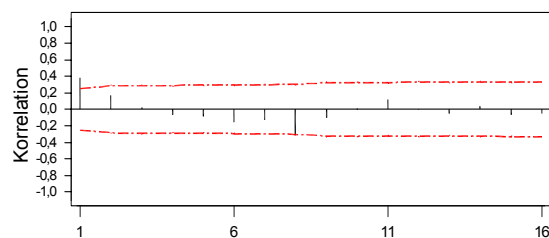
Lag	Korr	T	LBO	Lag	Korr	T	LBO	Lag	Korr	T	LBO
1	0,32	2,53	6,72	8	-0,02	-0,14	10,49	15	-0,05	-0,34	16,37
2	0,04	0,32	6,85	9	-0,09	-0,66	11,17				
3	0,00	0,01	6,85	10	-0,18	-1,27	13,81				
4	0,01	0,06	6,85	11	-0,11	-0,76	14,83				
5	-0,13	-0,95	8,07	12	-0,03	-0,22	14,91				
6	-0,17	-1,21	10,15	13	-0,05	-0,34	15,12				
7	0,07	0,45	10,46	14	-0,11	-0,73	16,14				

Diagram 4.4: Autokorelationsfunktion för export



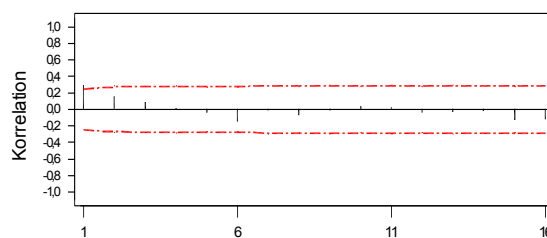
Lag	Korr	T	LBO	Lag	Korr	T	LBO	Lag	Korr	T	LBO
1	0,40	3,22	10,87	8	0,19	1,11	29,82	15	0,06	0,36	34,52
2	-0,01	-0,06	10,87	9	0,19	1,10	32,52	16	0,00	0,01	34,52
3	-0,18	-1,27	13,17	10	-0,02	-0,10	32,54				
4	-0,30	-2,00	19,32	11	-0,05	-0,29	32,74				
5	-0,32	-2,05	26,68	12	0,12	0,70	33,95				
6	-0,07	-0,43	27,06	13	0,04	0,26	34,12				
7	0,05	0,29	27,23	14	-0,03	-0,17	34,19				

Diagram 4.5: Autokorelationsfunktion för import



Lag	Korr	T	LBO	Lag	Korr	T	LBO	Lag	Korr	T	LBO
1	0,38	3,05	9,77	8	-0,32	-2,13	23,72	15	-0,07	-0,44	26,47
2	0,17	1,22	11,80	9	-0,11	-0,67	24,63	16	-0,05	-0,30	26,69
3	0,02	0,15	11,84	10	0,00	0,00	24,63				
4	-0,07	-0,47	12,16	11	0,12	0,72	25,72				
5	-0,09	-0,62	12,75	12	-0,01	-0,07	25,73				
6	-0,16	-1,06	14,51	13	-0,05	-0,31	25,94				
7	-0,14	-0,92	15,90	14	0,03	0,20	26,02				

Diagram 4.6: Autokorelationsfunktion för import av varor



Lag	Korr	T	LBO	Lag	Korr	T	LBO	Lag	Korr	T	LBO
1	0,29	2,36	5,82	8	-0,08	-0,54	10,20	15	-0,14	-0,95	12,21
2	0,16	1,15	7,48	9	-0,01	-0,09	10,21	16	-0,12	-0,85	13,55
3	0,08	0,57	7,91	10	0,03	0,24	10,31				
4	0,00	0,03	7,91	11	0,02	0,15	10,34				
5	-0,05	-0,35	8,08	12	-0,04	-0,29	10,49				
6	-0,15	-1,08	9,74	13	-0,03	-0,24	10,58				
7	-0,02	-0,13	9,76	14	-0,02	-0,16	10,62				

4.2 Modellsnittningar

I *Tabell 4.2* presenteras de modeller som utifrån tidigare nämnda kriterier (se *Avsnitt 2.3*) bedöms vara bäst för respektive variabel. De diagnostiska testerna kompletteras med en granskning av autokorrelationsfunktioner för residualerna (se *Appendix 1*) samt normalfördelningsdiagram som inte presenteras i uppsatsen.

Tabell 4.2
Modellsnittningar och diagnostik

Variabel	Modell	Param.	p-värde	Ljung-Box Q*	Ljung-Box Q* ²	JB ^b	A ²
				p-värde ^a	p-värde ^a	p-värde	p-värde
Privat konsumtion	(0,0,1)	$\delta = 0,286$	0,002	0,600	0,117	1,056	0,572
		$\theta_1 = -0,458$	0,000	0,759	0,300	0,590	
Offentlig konsumtion	(0,0,1)	$\delta = 0,761$	0,001	0,525	0,250	0,457	0,902
		$\theta_1 = -0,349$	0,005	0,202	0,278	0,796	
Investeringar	(1,0,0)	$\phi_1 = 0,359$	0,004	0,535	0,648	2,846	0,387
				0,697	0,939	0,241	
Export	(1,0,0)(1,0,0)	$\delta = 0,882$	0,000	0,124	0,911	0,554	0,931
		$\phi_1 = 0,377$	0,003	0,131	0,952	0,758	
		$\Phi_1 = 0,224$	0,088				
Import	(1,0,0)	$\delta = 0,696$	0,000	0,868	0,730	1,313	0,072
		$\phi_1 = 0,392$	0,001	0,335	0,548	0,519	
Import av varor	(1,0,0)	$\delta = 0,52$	0,000	0,690	0,312	1,400	0,059
		$\phi_1 = 0,296$	0,017	0,974	0,509	0,497	

a: Första raden anger p-värden för $m=6$ och den andra för $m=12$.

b: Första raden anger JB-värdet och den andra motsvarande p-värde.

Tabell 4.2 visar att samtliga skattade parametrar är signifikanta på femprocents nivå, frånsett säsongparametern för export, som dock är signifikant på tioprocentnivån. Vidare innehåller samtliga modeller (förutom investeringar) en signifikant konstant.

Ljung-Box testet visar att residualerna för samtliga modeller inte är autokorrelerade. Autokorrelationsfunktionerna för residualerna ger samma indikation, förutom för variabeln import. Där är residualerna signifikant autokorrelerade med åtta kvartals tidsförskjutning. Denna autokorrelation för åtta kvartals tidsförskjutning är också synlig i *Diagram 4.5*. En modellering av detta anser vi inte vara meningsfull eftersom ett beroende för åtta kvartal bakåt är av liten betydelse. Autokorrelation i residualerna för längre tidsförskjutningar än åtta kvartal, dvs. att ett beroende för två säsonger (2×4), förekommer för vissa modeller. Dessa bortses från av orsaker nämnda ovan. Ljung-Box testet för de kvadrerade residualerna (Ljung-Box Q*²) visar att vi inte kan förkasta antagandet om linjäritet i serierna för de skattade modellerna. Däremot kommer serien för offentlig konsumtion visa på icke-linjäritet om vi inte korregerar serien för extremvärden.

Jarque-Bera testet visar att modellantagandet gällande normalfördelade residualer är uppfyllt för samtliga modeller. Eftersom testet främst är avsett för stora stickprov har vi genomfört ytterligare ett normalfördelningstest, Anderson-Darling. Normalfördelningsantagandet kan med detta test inte förkastas på femprocentnivån, även om p-värdet för variablerna import och import av varor ligger nära förkastningsgränsen. Detta leder sammantaget till att samtliga modeller förefaller vara adekvata, möjligen med undantag av modellerna för import och import av varor.

4.3 Vad betyder resultaten?

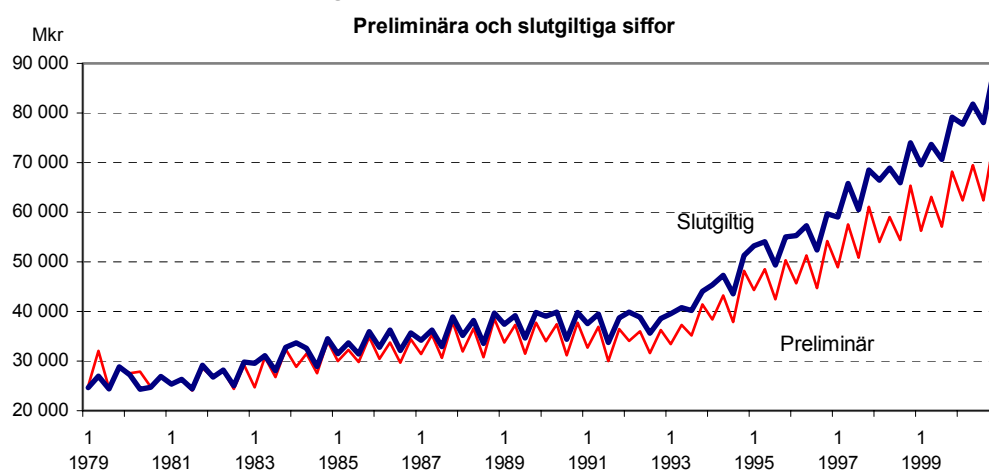
Revideringarnas autokorrelation för en periods tidsförskjutning kan åtminstone delvis bero på att den preliminära siffrans fel kan jämföras med felet i en två-steps prognos, eftersom jämförelsetalet för den preliminära tillväxtsiffran ännu inte är färdigreviderad. Enligt Box & Jenkins (1976, s.129) blir tvåstegsprognosers fel i allmänhet korrelerade.

Variablerna privat och offentlig konsumtion förklaras bäst med en MA(1)-modell. Det betyder att ett aktuellt kvartals revidering (y_t) beror, utöver kvartalets slumpmässiga avvikelse (a_t), på föregående kvartals slumpmässiga avvikelse (a_{t-1}). Vi har, trots försök, svårt att intuitivt tolka denna modell och vad detta beroende skulle vara orsakat av. Ett försök till tolkning är att modellen korrigerar för felet direkt. Felet ligger alltså inte kvar och "släpar" flera perioder framåt, utan nästa period (y_{t+1}) är det en ny slumpmässig avvikelse (a_t) som, förutom det slumpmässiga felet i perioden $t+1$, påverkar revideringen. Således förefaller producenterna av statistiken släpa efter i sitt beteende, men modellen korrigerar för detta irrationella beteende.

För investeringar, import och import av varor har vi identifierat en AR(1)-modell. Vi tolkar detta som att det existerar en tröghet i tidsserierna för dessa variabler som skulle kunna bero på irrationellt beteende vid beräkning och sammanställning av de preliminära siffrorna. Om vi jämför med MA(1)-modellen så ligger felet här kvar och släpar flera perioder framåt, fastän det (om man bortser från det slumpmässiga felet) hela tiden minskar.

Exportrevideringar är den enda variabel som visar på ett säsongberoende (som dock inte är rejält signifikant) och den lämpligaste modellen är en AR(1) med en AR-säsong av första ordningen. Det tyder på att den preliminära seriens säsongprofil skiljer sig systematiskt från den slutgiltiga. En ännu ej publicerad studie av Öller m.fl. visar mer tydligt på denna skillnad i säsongskomponenten för export. *Diagram 4.7* nedan är en rekonstruktion av de kvartalsvisa procentuella förändringarna för respektive serie i diagrammet. Detta innebär att vissa fel i form av avrundningar kan förekomma. Diagrammet visar att säsongmönstret för de preliminära siffrorna skiljer sig åt, i synnerhet efter år 1992. Detta skulle kunna betyda att revideringarna följer ett säsongmönster.

Diagram 4.7: Exportnivåer 1979 - 2000



Källa: Öller m.fl., SCB

5 Diskussion och slutsatser

I uppsatsen visar vi att revideringarna för privat och offentlig konsumtion, investeringar, export, import och import av varor är autokorrelerade och att det går att skatta modeller för dessa variabler. Därmed har vi kunnat identifiera ytterligare en typ av systematiska fel i BNP-revideringarna jämfört med studien av Öller & Hansson (2002). Positivt är att den totala BNP-serien inte är autokorrelerad. Att serien för privat konsumtion, som är ett av de större aggregaten, är autokorrelerad utgör dock ett problem som är värt en närmare granskning.

Vi kan också notera att de preliminära siffrorna skulle kunna förbättras avsevärt för flera variabler genom att korrigera för bias och, förefaller det, för systematiska fel i form av autokorrelation i serierna. En konsekvens av hur revideringarna beräknas är dock att de skattade modellerna i praktiken inte kommer att kunna användas för förbättring av de preliminära siffrorna. Detta till följd av att de värden (slutgiltiga siffror) som behövs för att förbättra de preliminära siffrorna kommer först ett år efter att dessa behövs för prognosen. Modellerna kan således *inte* användas till att förbättra de preliminära siffrorna genom prognostisering av revideringar, men kan vara till hjälp vid identifiering av de variabler där statistikproduktionen är i störst behov av översyn. Möjligen skulle man kunna förbättra den en gång reviderade siffran för året innan.

Litteraturförteckning

Bowerman, B. L. och O'Connell, R. T. (1993), *Forecasting and Time Series an Applied Approach*, tredje upplagan., Belmont, Duxbury Press.

Box, G. E. P. och Jenkins, G. M. (1976), *Time Series Analysis: Forecasting and Control*, reviderad upplaga, San Francisco, Holden-Day.

Enders, W. (1995), *Applied Econometric Time Series*, New York, John Wiley & Sons.

Fiorentini, G. och Maravall, A. (1996), "Unobserved Components in ARCH Models: An Application to Seasonal Adjustment", *Journal of Forecasting*, Vol 15, s.175–201.

Franses, P. H. (1996), *Periodicity and Stochastic Trends in Economic Time Series*, Oxford, Oxford University Press.

Gujarati, D. N. (1999), *Essentials of Econometrics*, andra upplagan., Singapore, McGraw-Hill.

Maravall, A. (1983), "An Application of Non-Linear Time Series Forecasting", *Journal of Business and Economic Statistics*, Vol 1., nr 1., s. 66–74.

SOU 2002:118, *Economic Statistics in Twelve Countries*, Stockholm, Fritzes förlag.

Öller, L.-E. och Hansson, K.-G. (2002), *Revision of Swedish National Accounts 1980–1998 and an International Comparison*, *SOU 2002:118*, Stockholm, Fritzes förlag.

APPENDIX 1 – autokorrelationsfunktioner för residualer

I Diagram A1.1–A1.6 nedan presenteras autokorrelationsfunktioner för modellernas residualer som utnyttjas vid utvärderingen av de skattade ARIMA-modellerna.

Diagram A1.1: Autokorrelation för residualer (privat konsumtion)

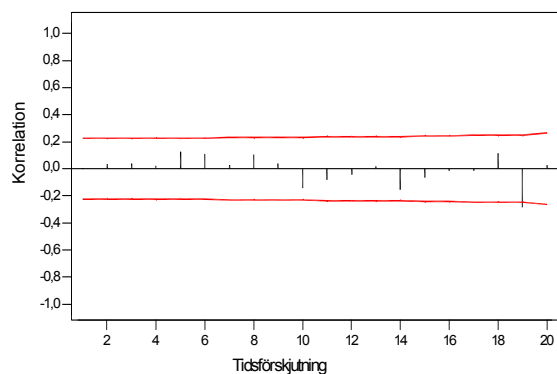


Diagram A1.2: Autokorrelation för residualer (offentlig konsumtion)

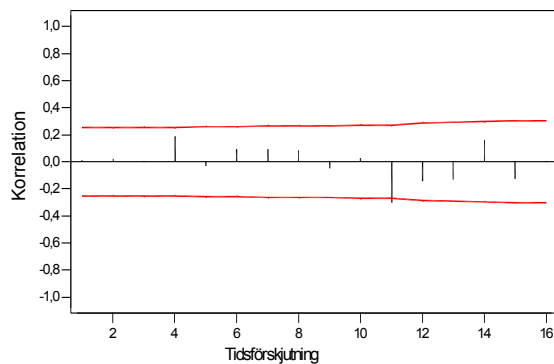


Diagram A1.3: Autokorrelation för residualer (investeringar)

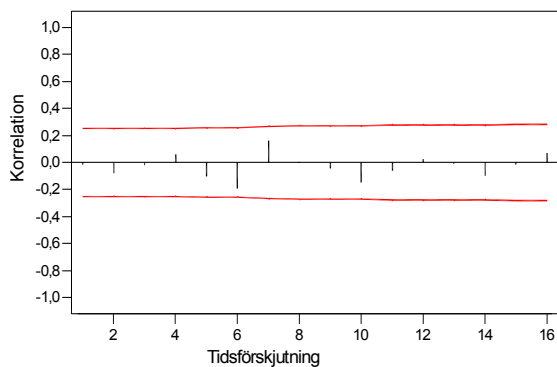


Diagram A1.4: Autokorrelation för residualer (export)

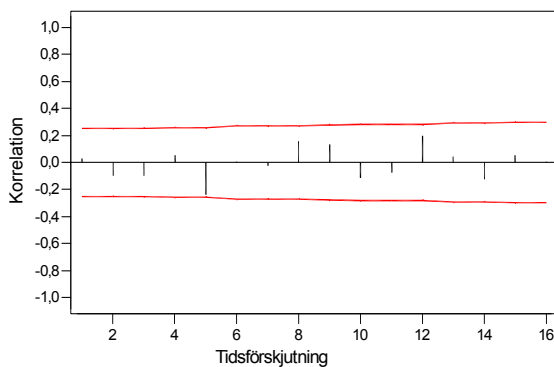


Diagram A1.5: Autokorrelation för residualer (import)

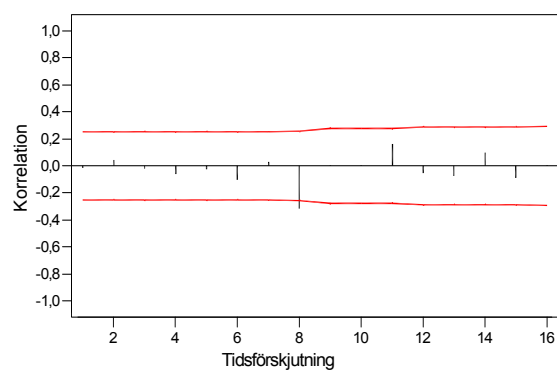
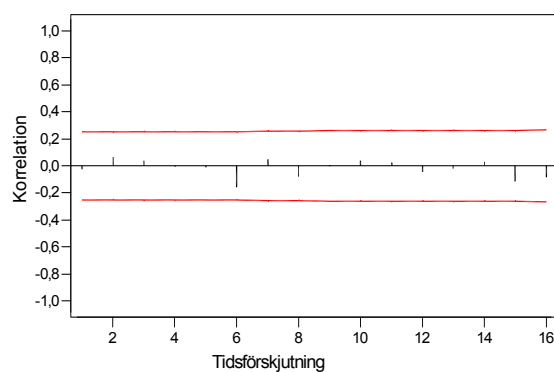


Diagram A1.6: Autokorrelation för residualer (import av varor)



- 2004:01 Hjälpverksamhet. Avrapportering av projektet Systematisk hantering av hjälpverksamhet
- 2004:02 Report from the Swedish Task Force on Time Series Analysis
- 2004:03 Minskad detaljeringsgrad i Sveriges officiella utrikeshandelsstatistik
- 2004:04 Finansiellt sparande i den svenska ekonomin. Utredning av skillnaderna i finansiellt sparande Nationalräkenskaper, NR – Finansräkenskaper, FiR
Bakgrund – jämförelser – analys
- 2004:05 Designutredning för KPI: Effektiv allokering av urvalet för prismätningarna i butiker och tjänsteställen. Examensarbete inom Matematisk statistik utfört på Statistiska centralbyrån i Stockholm

ISSN 1650-9447

Statistikpublikationer kan beställas från SCB, Publikationstjänsten, 701 89 ÖREBRO, e-post: publ@scb.se, telefon: 019-17 68 00, fax: 019-17 64 44. De kan också köpas genom bokhandeln eller direkt hos SCB, Karlavägen 100 i Stockholm. Aktuell publicering redovisas på vår webbplats (www.scb.se). Ytterligare hjälp ges av Bibliotek och information, e-post: information@scb.se, telefon: 08-506 948 01, fax: 08-506 948 99.

www.scb.se