

Ekonomisk statistik Economic statistics

Masterkurs

Daniel Thorburn

Höstterminen 2009

Stockholms Universitet

Innehåll

1. Introduktion
 1. Lite historik
 2. Indexteori
 3. Nationalräkenskaper
 4. Klassifikation, standarder
2. Tidsserier
 1. Länkning
 2. Säsongrensning
 3. Andra tidsaspekter
3. Urval och estimation vid företagsdata
 1. Ramar
 2. Varierande urvalssannolikheter π_{ps}
 3. Samordnade, roterande urval
 4. Fasta slumpstal
 5. Outliers
4. Datainsamling från företag
 1. Allmänt
 2. Modeller för svarsprocessen
5. Editing (Granskning och rättning)
 1. Introduktion till granskning
 2. Missing values
 1. Modeller
 2. Imputering
 3. Reweighting
6. Avslutning
 1. Kvalitet mm

1.1 Kort historik

- Statistik har gamla rötter - Fjärde Mosebok, Numerii, handlar till stor del om folkräkning
- Det finns också gamla dokument/inskrifter från Egyptisk tid med folkräkning för att veta hur många arbetare som varje distrikt skulle sända till pyramidbyggena
- Den äldsta ekonomiska statistiken (som jag känner till) är ett mellanting av nationalräkenskaper och ekonomisk bokföring för det gamla Sumerien (på kilskriftstavlor i lera)
- Axel Oxenstierna tog fram de första bevarade handelsbalanserna i Sverige 1637
- Kommerskollegium (Emanuel Swedenborg) fick 1723 i uppdrag att kontinuerligt följa handelsbalansen. Deras ”persedelextrakt” är förmodligen Sveriges äldsta statistiska blanketter

1.2 Introduktion till indexteori

Daniel Thorburn
Ekonomisk statistik
Höstterminen 2009

Historik för konsumentprisindex

- Biskop Fleetwood, Cronicon Preciosum, 1707 (Vad motsvarade år 1707 inkomsten £5 år 1450
Svar £28, baserat på priser för Säd (£30), Kött (£30), Dryck (>£25) och Tyg (<£25))
- Joseph Lowe 1822, första teoretiska diskussionen, för in kvantiteter som vikter
- Irving Fisher, 1922, En portalfigur
- Divisia, 1925, motiverar kedjeindex
- Malmquist (ca 1960) bör nämnas här eftersom han var professor vid denna institution (Andra viktiga nordbor är finländarna Törnquist och Vartia)

Ekonomisk teori.

- Kostnad för att uppnå viss nyttonivå, indifferenskurvor
- Efterfrågesidan konsumenterna indifferenta Utbudssidan producenterna indifferenta
- ”Indextest” Önskemål på indexar som inte alltid är så enkla att uppnå, t ex ”Om priser och kvantiteter är desamma år 1 och 3 skall det kedjade indexet bli 1 oberoende av priser och kvantiteter år 2” eller ”Om alla priser ökar med samma faktor c så skall också prisindex ändras med faktorn c ”
- Samband produktivitet, kvantiteter och priser. Relativt lätt att uppskatta BNP i kronor, men hur mycket har produktionen ökat i reala priser?

Olika indexar

– Några klassiska index

- Laspeyre kvantiteter vid tid 1, Paasche vid tid 2.
- Normalt ger Laspeyre högre värden. Varför?
- Men Laspeyre mer praktiskt. Varför?
- Problemet med för stor inflation är svårt. USA

– Lösningsförslag

- Fishers ideala index. Geometriskt medelvärde
- Andra indextyper ...

– Hedoniska

- Pris inte på produkter utan på egenskaper.
- T ex Linjär regression ger datorpris = $a + b \cdot \text{minneskapacitet} + c \cdot \text{klockfrekvens} + d \cdot \text{skärmtyp}$. Använd a,b,c och d som ingredienser

Olika indexar

- Fastighetsprisindex, konsumentprisindex, exportprisindex, lönekostnadsindex, ...
- Delindex som skor, kläder, bostäder, hopsummering (skor var speciellt aktuellt förra året. Varför?)
- PPP (purchasing power parity, jämförelse inte över tiden utan mellan länder)
- Levnadskostnadsindex var kostnader för en typfamilj medan KPI är kostnader för en genomsnittsfamilj ("plutokratiskt index")

PROs mätning

- Lite annat syfte
- I korgen finns bara dagligvaror
- Valda subjektivt som lämpliga varor som ”normala” fattiga pensionärer köper (jfr levnadskostnadsindex)
- Bara vanliga livsmedelsaffärer
- Fast korg dvs Laspeyre-typ med dess bias

Penningvärdet i Sverige

Myrdal-Bouvins index 1858–1914, Levnadskostnadsindex

1914-1947, KPI 1947-2008

Genomsnittlig inflation 2,7%

1858	100 rd bco	1938	200 kr
1868	108 rd bco	1948	310 kr
1878	108 rd bco	1958	475 kr
1888	92 kr	1968	675 kr
1898	100 kr	1978	1 465 kr
1908	115 kr	1988	3 150 kr
1918	295 kr	1998	4 580 kr
1928	210 kr	2008	5 430 kr

Datainsamling - teori

– Sampling

- Vanligen – drag ett urval på n enheter ur en population med N individer
- Korsklassificerad sampling – drag ett urval varor (korg) och ett urval butiker och studera skärningen (proportionellt mot konsumtion resp omsättning)

– Varuvikter

- HUT hushållets utgifter – en kassaboksstudie där ett antal hushåll för detaljerade kassaböcker. Svårt motivera och därmed stort bortfall (drygt 50%)

– Korgbyte

- Görs varje årsskifte, vilket leder till länkningsproblem (Jfr skor ovan)

Datainsamling

– några speciella problem

- Kvalitetskorrektioner: T ex vad i prisändring i en bil är kvalitetsändring och vad är prishöjning?
- Extrapris: T ex hälften av allt kaffe säljs till extrapris – men om representantvaran t ex är Gevalia halvkilo, kanske den bara säljs till extrapris tre veckor om året
- Rea: T ex nästan alla kläder blir billigare. De reas ut men detta skall inte spegla prisutvecklingen
- Boendekomponenten:
 - Vad är konsumtion och vad är sparande vid egnahem?
 - När riksbanken höjer räntan för att få ner inflationen ökar KPI
- Nya varor: T ex mobiler. Skall man följa priset på telefonsamtal eller bör man inte jämföra fast och mobiltelefoni. För några år sedan en debatt om priset på ljus (elljus)
- Subventionerade varor: Hur skall priset på t ex sjukvård räknas eller barnomsorg?
- ...

Fler problem

- Om en vara försvinner räknas den inte i KPI. Men om den fortfarande är åtkomlig men till tredubbla priset ökar prisindex mycket
- Hur ta hänsyn till skift i kanaler. När små närbutiker läggs ner och ersätts av billiga stormarknader utanför city så påverkar det idag inte index. Men för en del innebär det en kraftig prishöjning eftersom bil/transportkostnaden tillkommer.

Publicering

Konsumentprisindex fastställs i en index-nämnd och sedan gäller den publicerade siffran.

- Anledning: Det finns så många indexreglerade avtal som är knutna till KPI.
 - Om man publicerade KPI på sedvanligt sätt KPI ligger i intervallet 4,5 % +/- 0,8 %, skulle det bli alltför många processer. (Prisklausul som ger rätt till omförhandling om KPI överstiger 5%).
 - Nu fick pensionärer och barnfamilj lite högre pension och bostadsbidrag pga felräkningen.
 - Riksbanken skall hålla inflationen runt 2 %. Tidigare höll man sig med en egen definition "Underliggande inflation", där man bl a räknade bort effekten av Riksbankens egna ränteändringar.

1.3 Introduktion till nationalräkenskaperna

Daniel Thorburn
Ekonomisk statistik
Höstterminen 2009

Nationalräkenskaper

Historik

- William Petty, 1665, Verbum Sapientia, Förgrundsfigur i politisk aritmetik, (syfte militära resurser)
- Simon Kusnetz 1934, första officiella nationalräkenskaper (i USA). (Pris till Nobels minne) Viktig del i det Keynesianska tankebygget efter finanskrisen på 30-talet. Grundläggande teori utvecklas i England under 1930- och 1940-talet.
- Richard Stone, 1947, och mannen bakom 1953 SNA
- I Sverige började Konjunkturinstitutet med nationalräkenskaper i början av 1950-talet
- 1968 SNA, nu var räkenskaperna ”helt färdigutvecklade” och det teoretiska intresset avtog bland (universitets-)forskare
- Nu har intresset ökat igen med t ex olika typer av satelliträkenskaper
- Historiskt har man vid ekonomisk historia i Lund (bl a Jörberg och Schön) skapat historiska nationalräkenskaper för Sverige från 1700-talet.

Lite teori

- BNP värdet av allt som produceras (ungefär lönekostnader + vinster); BNI är BNP justerad med utrikes transfereringar
- Ungefär: produceras med betalt arbete – t ex dagis men inte barnpassning inom hushållet eller åldringsvård på vårdhem men inte av släktingar. Mat på restaurang ja men inte i hemmet (om det inte var t ex betalda pigor som lagade maten)
- BNP till fasta eller löpande priser. Hur deflatera? T ex vi vet inget om produktiviteten hos offentliga sidan. Ökar produktiviteten hos t ex lärare, forskare, sjuksköterskor eller SCBs personal
- Tillgång – användning (tillgång är ungefär vad som produceras och användning vad som används/konsumeras). Skillnad investeringar/lager mm
- För att åstadkomma nationalräkenskaper görs en avstämning mellan produktions och användningssidan. (ca 400 produktbalanser). Input-outputtabeller.
- Dessutom finns en inkomstsida, som spelar mindre roll i Sverige. Dvs produceras som BNI ovan

Nationalräkenskaper

- Idén är att lägga samman det penningmässiga värdet på allt som produceras utan att lägga några moraliska aspekter på vad som produceras hur det produceras eller hur det används.
- Dvs ingen skillnad om det förstörs i krig eller naturkatastrofer eller om sparas till kommande generationer som byggnader.
- Ingen skillnad om produktionen förstör hälsa och miljö eller om det är hållbar produktion. Om man betalar extra för hållbar produktion eller bättre arbetsmiljö ökar BNP däremot eftersom det penningmässiga värdet ökar.
- Ingen skillnad på produktion av barnporr eller socialvård. (Även illegala aktiviteter skall i princip räknas)
- Inte ej betald produktion som värnpliktiga, oavlönat ideellt arbete tex i idrottsföreningar eller med välgörenhet. Studier räknas inte om man inte får betalt (som doktorandtjänst)
- Med tiden har fler och fler verksamheter inkluderats t ex egna förbättringsarbeten på egen villa eller bostadsrätt (Harmonisering)

Andra räken-skaper

- Sektorsräken-skaper (jordbruks-, bygg-, tjänste- osv)
- Regionala räken-skaper – EU har skapat ett system av regioner i Europa
- Satelliträken-skaper – T ex miljö, turism eller hälsa. Inte bara input i form av löner och vinster räknas utan också påverkan på miljö.
- Under kommunisttiden hade dessa länder (t ex Kina och Sovjet) ett annat nationalräken-skaps-system baserat på mervärdesbegreppet

1.4 Standarder

Daniel Thorburn
Ekonomisk statistik
Höstterminen 2009

Standarder

- SNI-2007 Svensk näringsgrensindelning (bygger på EU-standard NASE rev2)
 - uppbyggd enligt decimalsystemet
 - Text 30xxx är kontorsmaskiner och datorer 30010 tillverkning av kontorsmaskiner
 - 80xxx olika former av utbildning, text 80 410 trafikskolor

http://www.scb.se/Grupp/Foretagsregistret/_Dokument/070129kortversionSnisorterad2007.pdf

- SSYK svensk yrkesklassificering (1996)
- SPIN (CPA från bl a EU) Produktindelning efter näringsgren (2002)
- HS/KN Tullnomenklatur Intrastat varuklassificering (Harmonised Commodity Description and Coding System)
- SITC FNs varugrupper för utrikeshandel (2006)
- SUN-2000 Utbildningsnomenklatur

COICOP Classification of Individual Consumption According to Purpose

1. [01](#) - Food and non-alcoholic beverages
2. [02](#) - Alcoholic beverages, tobacco and narcotics
3. [03](#) - Clothing and footwear
4. [04](#) - Housing, water, electricity, gas and other fuels
5. [05](#) - Furnishings, household equipment and routine household maintenance
6. [06](#) - Health
7. [07](#) - Transport
8. [08](#) - Communication
9. [09](#) - Recreation and culture
10. [10](#) - Education
11. [11](#) - Restaurants and hotels
12. [12](#) - Miscellaneous goods and services
13. [13](#) - Individual consumption expenditure of non-profit institutions serving households
14. [14](#) - Individual consumption expenditure of general government

COFOG

- **Classification of the functions of the government (COFOG)**
 1. *General public services*
 2. *Defense*
 3. *Public order and safety*
 4. *Economic affairs*
 5. *Environmental protection*
 6. *Housing and community amenities*
 7. *Health*
 8. *Recreation, culture and religion*
 9. *Education*
 10. *Social protection*

Standarder

- Alla dessa är uppbyggda enligt decimalsystemet (SNI är en ren bransch-kod men t ex SSYK klassificerar både efter bransch och nivå)
- Men ett intressant problem är hur man skall utforma standarder. De skall vara heltäckande och entydiga, både lättkodade för den som samlar in data och meningsfulla för statistikanvändaren
- Dessutom skall de vara både stabila över tiden och rummet (Jämförelser mellan länder och mellan år) men samtidigt skall de vara följa utvecklingen. Både branscher, varor, yrken och utbildningar ändras med tiden och har olika mening i t ex västvärlden och utvecklingsländer.

Harmonisering

- Termen för att åstadkomma jämförbar statistik i EU/hela världen
- Gäller inte bara klassifikationsstandarder
- Jämföra t ex arbetslöshetsbegrepp, nationalräkenskaper, inflation
- Göra studier på samma sätt vid samma tidpunkt på året etc.
- Ibland blir Sverige tvingat att göra sämre undersökningar än vad vi skulle kunna (med våra register) bara för att få internationell jämförbarhet.

2.1 Tidsserier - Länkning

Daniel Thorburn
Ekonomisk statistik
Höstterminen 2009

Länkning

- När statistiken läggs om uppstår ofta språng, t ex. när
 - näringsgrensindelningen moderniseras
 - man byter insamlingssystem t ex från disketter till Internet
 - en cutoff-gräns ändras.
 - Man formulerar om en fråga
- Vid länkning försöker man räkna om serien bakåt som om det nya insamlingssystemet hade gällt. Bra för personer som arbetar med tidsserier.
- Samma problem kan uppstå vid ändringar i samhället.
 - Följa sjuklighetens utveckling i samhället, när antalet karensdagar ändras.
 - Vad händer med data över hushållens förmögenhet när förmögenhetsskatten avskaffas?
- Vid omläggningar anses länkningen inte alltid som en uppgift för statistikproducenten utan snarare tidsserieanalytikern. Men länkning vid metodskift är något för statistikproducenten.

- Vid omläggningar bör statistikern planera för länkning. Vid reformer i samhället bör politikerna planera så att reformen kan utvärderas.
- Det finns två typer av länkning (med mellanformer)
 - Mekanisk länkning (Den enda källa du har är serien själv och det gäller att uppskatta språnget)
 - Länkning med hjälp av speciella data
 - Specialundersökningar
 - Andra relevanta data
 - Gamla rådata, blanketter

Mekanisk länkning ?

Tänk er att man har en serie som ser ut så här
(Andel positiva till statliga investeringsgarantier %.)

År	1	2	3	4	5	6	7	8	9	10	11	12	13
	23	24	27	27	26	28	30	42	43	45	44	43	47

År 8 lades undersökningen om och frågan
formulerades om. Om man vill analysera
utvecklingen vill man ha en serie utan
språng. Hur skall man göra?

•

- **Svar 1 o 2:** Enklast är att sätta värdena år 7 och 8 lika och ändra värdena innan motsvarande (additivt resp multiplikativt).

År	1	2	3	4	5	6	7	8	9	10	11	12	13
Urspr.	23	24	27	27	26	28	30	42	43	45	44	43	47
add	35	36	39	39	38	40	42	42	43	45	44	43	47
Mult	32.2	33,6	37.8	37.8	36.4	39.2	42	42	43	45	44	43	47

Problem: Detta säger inget om skillnaden mellan år 7 och 8.

- **Svar 3:** (För enkelhets skull ges lösningen bara additivt)

Mer avancerat gör regression $y_t = a + b \cdot I(t < 8) + c \cdot t$ och justera med b

34,9 35,9 38,9 38,9 37,9 39,9 41.9 42 43 45 44 43 47

Andra modeller kan användas. Regression mot en logistisk linje hade förmodligen varit ännu bättre (grundtips: om andel så logistisk regression).

$$p_t = \exp(a + b \cdot I(t < 8) + c \cdot t) / (1 + \exp(a + b \cdot I(t < 8) + c \cdot t))$$

$$y_t = p_t + \varepsilon_t; \quad \text{Var}(y_t) = d \cdot p_t \cdot (1 - p_t)$$

•

Problem: Man vill kunna göra länkningen redan första året efter tidsseriebrottet (dvs år 8).

- **Svar 4:** Använd samma modell redan år 8 och håll sedan modellen fix
34,6 35,6 38,6 38,6 37,6 39,6 41,6 42 43 45 44 43 47

Ger lite större osäkerhet än svar 3.

- **Svar 5:** Ännu mer komplicerad modeller kan användas t ex en ARIMA-process eller en dynamisk modell (Kalmanfilter). T.ex
 - $X_t = a * X_{t-1} + \varepsilon_t$ X är en latent serie som mäts med slumpfel:
 - $Y_t = X_t + \eta_t$ före brott; $Y_t = X_t + b + \eta_t$ efter brott;
 - Y + Predikterat värde av b redovisas före brottet
 - Den dynamiska modellen ger
 - 34,9 35,9 38,9 38,9 37,9 39,9 41,9 42 43 45 44 43 47

- Mekaniska modeller som 3-5 ger möjlighet att ge ett osäkerhetsintervall för justeringens storlek. I så fall bör man ha en relativt flexibel modell som den dynamiska modellen.
 - Medelfel i språnguppskattningen i vårt exempel var 1,8 i svar 5.
 - Osäkerhet i språnget blir +/-3,6.
- Det är dock mycket ovanligt att man anger osäkerheten i uppskattningen av länkningsfelet, när den görs mekaniskt, som här.
- Men i allmänhet försöker man välja så enkla modeller som möjligt.

Hjälpsier

- I diskussionen ovan antogs att man inte hade någon annan kunskap. Ofta har man det.
- T ex kan man ha tillgång till serier för en eller flera andra relaterade variabler (I vårt fall t ex branschindelning, partisympatier eller investeringsvilja). Då kan dessa variabler också användas i regressionsmodellerna. (eller göra om ARIMA-modellerna till VARIMA)

- Till exempel gör regressionen

$$y_t = a + b \cdot I(t < 8) + c \cdot t + d \cdot x_t + e \cdot y_{t-1} + a_t$$

där x är hjälpserien och justera sedan med skattat b

- Ibland kan denna typ av modeller leda till att man inte justerar alla åren bakåt i tiden lika mycket. Man kan t ex komma fram till att alla moderater skall justeras med 7%, övriga borgerliga med 10% och övriga med 12%. Sedan justerar man bakåt med de tidigare observerade partisympatierna. Eller efter branschernas historiska andel av totalen.

Länkning med hjälp av tilläggsstudier

Allt tidigare var s.k. mekaniska länkningsmetoder. Om möjligt bör man göra specialstudier vid alla större omläggningar.

- En vanlig rekommendation är att om det är möjligt göra undersökningen på båda sätten parallellt vid ett tillfälle
- T ex När ULF (undersökningen om levnadsförhållanden) gick över från besök till telefon-undersökning som bas gjorde man under övergångsåret halva urvalet besök och halva per telefon. (Delarna valdes slumpmässigt, men varje intervjuare fick bara använda en metod, ny eller gammal).
- Men vanligare är att man gör mindre specialundersökningar av olika typer.
- Ibland har man tom gjort experiment i förväg för att finna en bra ny uppläggning. Då kan man i underlaget för beslutet om omläggning även ha tillräckliga data för att uppskatta språngets storlek.

Flera referenstidpunkter för länkning

- Om man bara har en tidpunkt för jämförelse måste alla ändringar göras med samma kvantitet bakåt i tiden (additivt eller multiplikativt)
- Men med flera mätningar kan man modellera en gradvis övergång

År	1	2	3	4	5	6	7	8	9	10	11	12	13
Urspr.	23	24	27	27	26	28	30	42	43	45	44	43	47

Man har en engångsundersökning år tre, där både den nya och den gamla frågan ställdes parallellt men till en delvis annan population.

Då blev värdena 10% lägre med nya frågan.

År 8 var den 30% lägre (enligt svar 3). Då räknar man om och tror att relationen steg från 10 % till 30 % dvs med 4 % mer varje år dvs man räknar om med 2, 6, 10, 14, 18, 22, 26, 30 % och får

Länkat: 23 25 30 31 31 34 38 42 43 45 44 43 47

Länkning med underlag

- ”Inbäddade experiment” är en speciell typ av studier. De görs som en del i den normala statistikproduktionen
- När uppskattningen av hoppet görs med specialundersökningar är det vanligt med feluppskattningar
- När större omläggningar av klassificeringar görs, t ex av SNI-kod försöker man
 - koda uppgifterna på båda sätten under en period eller
 - gå igenom gamla statistikunderlag (blanketter) och koda om företagen enligt den nya mallen och därmed få jämförbara data för tidigare år.

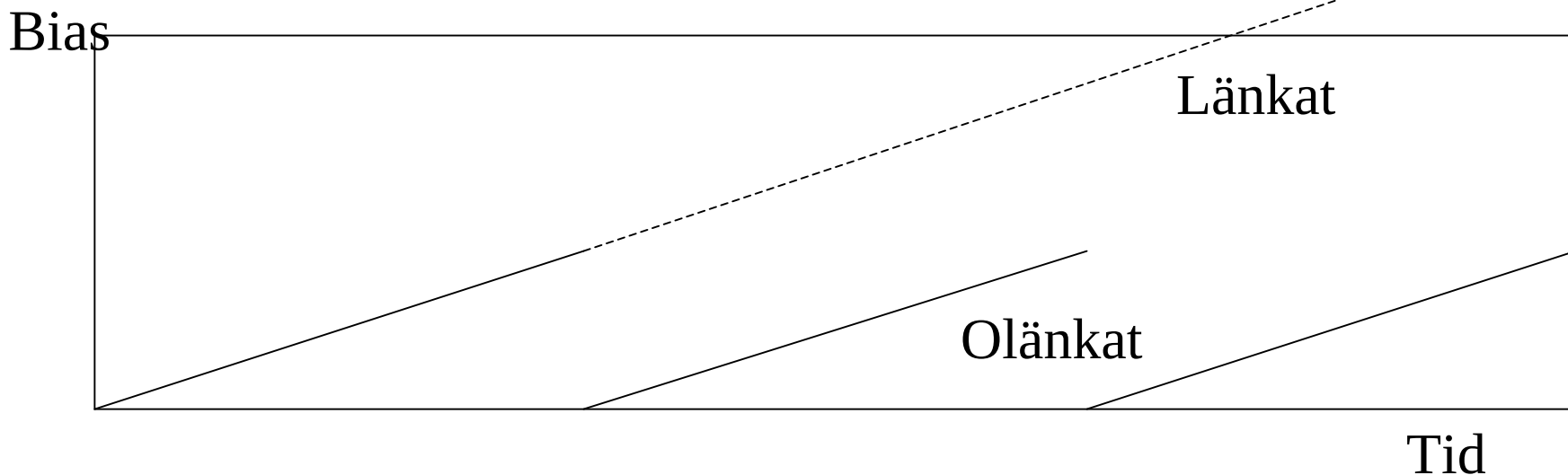
- Dessa metoder ger en uppskattning av språnget - men inte över hur man skall ändra längre bakåt
- Vid bl a omklassificering, t ex omläggning av SNI, kanske man vet summan av två grupper men det tillkommer en uppdelning.
 - T ex när Dator och IT-konsulter bröts ur kategorin annan konsultverksamhet. Då visste man att den inte under de senaste tio åren haft samma andel av den större kategorin.
- Då gäller det att rekonstruera en uppdelning bakåt. Om man inte har data lägger man ibland in en subjektiv gissning. T ex andel 0 tio år tidigare och sedan en linjär ökning av andelen till den första observerade nivån. Eller samma andel hela tiden bakåt (Dumt i just detta fall)
- Ett problem med länkning är att serien blir utjämnad. (Inte för producenten men för ekonometrikern, som använder data). Ekonometrikern kommer att underskatta slumptermens storlek när han sedan använder t ex vid ARIMA-modeller. Det betyder bl a att hans prognoser blir för säkra.

Flera serier – direkt/indirekt länkning

- Vid flera omlagda serier som skall summera till en total kan man
 - länka delserierna och sedan summera
 - länka totalerna och delserierna var för sig
 - Då får man inte konsistens
 - länka totalen och sedan länka delserierna
 - med bivillkor att länkningen skall vara konsistent
- Man talar om indirekt och direkt länkning
- Vilket som bör väljas beror på vad man tror om stabiliteten
 - t ex om man har två serier varav en lagts om och inte den andra och man skall göra mekanisk länkning är det alltid bättre att länka den serie som förändrats
 - Men om man har en total som beror av två delar och omläggningen främst består av omklassificeringar så bör man länka totalen först.

- Även vid flera serier är det ibland lämpligt att använda modeller. Vi diskuterade univariata serier - men man kan lika gärna använda multivariata modeller.
- Om man gör specialstudier av länkningsfelet för några delserier är det naturligt att länka dessa delserier först. Det är inte givet hur studierna skall användas när man länkar de andra delserierna eller totalen. Men ofta kan man utnyttja informationen från de studerade delserierna även för övriga.
(det går att länka några delserier och summan först och sedan anpassa övriga serier som en eller flera resttermer)

- I praktiken görs ofta små ändringar i statistikinsamlingen. I princip skulle de kvalificera för en länkning, men om effekten är liten struntar man ofta i det. Felet i uppskattningen av språnget blir större än språnget själv.
- Ofta kan man säga att när man gjort en stor omläggning innehåller studien bara små fel. Sedan förändrar sig verkligheten och studiens relevans blir sämre och sämre.
- Efter tio år, säg, blir man tvungen att göra en ny stor omläggning. Om man bara justerar för en långsam biasökning riskerar man att på lång sikt hamna fel



- Om man länkat en dataserie, skall den som använder serien alltid kunna läsa om vad som hänt. Om orsaken till språnget och hur det eliminerats.
- Dokumentera vad du gör. Detta är en allmän regel, som man alltid säger, men många slarvar med.
- Användarna kan ha mer kunskaper än du och vilja göra en egen länkning/säsongrensning ...

3.2 Tidsserier - Säsongrensning

Daniel Thorburn
Ekonomisk statistik
Höstterminen 2009

Komponentuppdelning

- Serien $Y =$
 - Trend T (långsiktig utveckling) $T +$
 - ”Cykler” C (Övergående avvikelser från trend som inte är säsong eller begränsade till en period t ex konjunktur. inte säkert cykliska) $+$
 - Säsong S (T ex årstidseffekter, julhandeln, semesterförläggning, när studerande/värnpliktiga blir färdiga och rycker ut...) $+$
 - Kalendereffekter K (månadslängd, påskens placering, antal icke-arbetsdagar etc) $+$
 - ”Slumpfel” E (periodeffekter, mätfel + väder, strejker etc)
- Additivt - multiplikativt

Enkla metoder

- Jämförelse med samma period (kvartal/månad) föregående år
- Glidande medelvärden (t ex summa av senaste period 12 månader eller flera period)
- Säsong + Slump = Värdet minus glidande medelvärde
- Säsong* = (vägt) medelvärde av de senaste säsong + slump för samma månad några år bakåt

Enkelt exempel: Modell $Y = T + S + E$

- Gör 12 månaders medelvärde Sept 2008
 $= (\frac{1}{2}\text{mar08} + \text{apr08} + \text{maj08} + \dots + \text{feb09} + \frac{1}{2}\text{mar09}) / 12$
- Detta ger T^* och $S^* + E^*$ skattas som $Y - T^*$
- Femårs glidande medelvärde ger säsong Sep06 =
 $(\text{Sep04} + \text{sep05} + \text{sep06} + \text{sep07} + \text{sep08}) / 5$
- Detta ger S^* och därmed också E^*
- Andra viktade medelvärden kan användas
- Här ser man problemet. Man får ingen bra rensning förrän efter $2\frac{1}{2}$ år
- Finns flera typer av lösningar. Vi skall ta upp tre olika ideer mycket kort

Exponentiell utjämningsgruppen

(Medelvärdena görs bara bakåt i tiden)

- Exponentiell utjämning $Y^*_t = \alpha Y_t + (1-\alpha) Y^*_{t-1}$
- Exponentiell utjämning med säsong

$$Y^*_t = \alpha(Y_t - S_{t-p}) + (1-\alpha) Y^*_{t-1}$$

$$S_t = \beta (Y_t - Y^*_t) + (1-\beta) S_{t-p}$$

- Holt-Winters metod utan säsong (exponentiell med trend)

$$Y^*_t = \alpha Y_t + (1-\alpha)(Y^*_{t-1} + T_{t-1})$$

$$T^*_t = \beta(Y^*_t - Y^*_{t-1}) + (1-\beta)T^*_{t-1}$$

- Holt-Winters metod med säsong ...

SEATS-Tramo/X12-Arima-gruppen

(Använd ARIMA-modeller för den rensade serien)

- Avancerade programpaket. SCB använder idag SEATS-Tramo
- Svart låda
- I princip bygger de på att säsong + slump är värdet – vägt glidande medelvärde för en period p perioder bakåt och framåt
- Men framåt vet man inte.
- I princip anpassar man då en ARIMA-process och predikterar framtiden. Därefter används de predikterade värdena som verkliga.
- Valet av tidsseriemodell mycket viktigt – sker ad hoc – genom att titta på resultatet

Dynamiska modeller – (mer likt vanlig teoretisk statistik)

- Bygg modell för allt dvs även den bakomliggande latent ”parametrar” t ex
 - Säsongmönster $S_t = S_{t-1} + \varepsilon_{st}$ $\text{Cov}(\varepsilon_{st}) = \Sigma$
där $\sum_i \sigma_{ij} = 0$ (samma månad alltid först)
 - Trend $\Delta T_t = \Delta T_{t-1} + \varepsilon_{\Delta T_t}$
 - ”Cykel” $C_t = \beta C_{t-1} + (1-\beta) C_{t-2} + \varepsilon_{\theta t}$
(Stationär AR(2))
 - $Y = C_t + T_t + S_t + \varepsilon_{Yt}$
- Skatta parametrar och prediktera ingående variabler simultant

Direkt – indirekt rensning

- Ofta har man många komponenter och summan $BNP = \text{summa}(\text{alla sektorer})$. Skall man då
 - säsongrensa totalen för sig och sedan justera komponenterna (direkt säsongrensning)
 - eller rensa varje komponent först och ange summan av de säsongrensade termerna (indirekt säsongrensning)
- Ingen enighet om svaret.
 - Utjämnarna oftast direkt
 - ARIMA-folket ofta direkt om ungefär liknande ARIMA-modell – indirekt om helt olika modeller
- Normalt svårare vid multiplikativa (del-)serier

Kriterier för bra rensning

- Idempotent (om man säsongrensar en rensad serie skall inget hända). Gäller normalt inte exponentiell utjämning, men däremot i allmänhet metoder baserade på dynamiska modeller. ARIMA-folket där emellan
- Små svängningar Jämför t ex $\sum (Y^*_t - Y^*_{t-1})^2$ sedan säsong och periodeffekter tagits bort. Plotta.
- Små förändringar när nya observationer tillkommer. Det säsongrensade värdet för innevarande tidpunkt skall inte ändras när data för nästa period kommer.

- Dokumentera alltid den säsongrensning som gjorts

2.3 Andra tidsaspekter

Daniel Thorburn
Ekonomisk statistik
Höstterminen 2009

Speciell tidsserieproblematik

- Många tidsserier har en specifik struktur på mätfelen. Det kan vid analysen vara bra att känna till det. Annars kan man hitta felaktiga samband med t ex ARIMA-modeller
- Ex. Arbetskraftsstatistiken. Varje månad är 1/8 av urvalet nytt och nyvalda återkommer sedan var tredje månad i två år.
Korrelationsfunktionen för mätfel är ungefär $r(t) =$
 - 1 om $t=0$
 - $\sim (24-t)/24 * r^t$ om t jämnt delbart med 3 och mindre än 25 (r varierar beroende på vilken variabel men kan var t ex 0,95)
 - 0 om t inte är delbart med tre eller större än 23
- Med ARIMA-analys hittar man en period runt tre månader
- Konsumentprisindex gör ett korgbyte vid varje årsskifte, som man också kan hitta om man letar efter det
- Säsongrensning och länkning åstadkommer också problem. En säsongrensad serie är normal jämnare än vad den orensade serien är.

Longitudinella studier

- Longitudinella studier är ett samlingsnamn för studier där man följer samma enheter över tiden
- Motsatsen är tvärsnittsstudier
- Studera t ex de företag som får regionalstöd och se vad som händer med dem, jämfört med dem som inte fått stöd. Då kan man också ta hänsyn till typ av stöd och dess storlek.
- Eller undersök vilka företag som växer flera år i rad

Inspektionsparadoxen mm

- Tänk er att ni vill studera utvecklingen på svenska börsföretag. Ni väljer ut de tio största och studerar hur omsättningen förändras under en period av 5 år
- Då kommer ni förmodligen att underskatta utvecklingen. Varför?

Inspektionsparadoxen mm

- Tänk er att ni vill studera utvecklingen på svenska börsföretag. Ni väljer ut de tio största och studerar hur omsättningen förändras under en period av 5 år
- Då kommer ni förmodligen att underskatta utvecklingen. Varför?
- Rätt vore att jämföra de tio största år 0 med de 10 största år 5. Vissa företag ökar mycket och några lite. De största kan bara tappa placeringar.

Inspektionsparadoxen mm

- Man gör en specialundersökning under september. De som svarar att de gjort en miljöinvestering väljs ut till en specialundersökning där man frågar om alla miljöinvesteringar under perioden sedan september året innan
- Varför blir detta en skev skattning av miljöinvesteringsverksamheten?

Redovisningsårets längd

- Vissa företag har bara 6 månaders räkenskapsår andra har 18. Vissa har juli-juni; andra har jan-dec
- Det kan vara frestande att räkna om åren med en faktor 2 resp $2/3$. Det görs ofta, men kan ge stora fel i branscher med stora säsongskillnader.
- Bättre är att i första fallet för ett visst år ta de månader som ligger det året. T ex med två räkenskapsår jan-aug, sept-aug skattas årssiffrorna med hela första perioden + $1/3$ av andra året. Då får man additivitet mellan åren.
- Om man har möjlighet kan man först säsongrensa – men det har man sällan
- Ett annat problem är att förkortade eller förlängda räkenskapsår ofta hänger samman med strukturförändringar. Då passar man ofta på att göra ovanligt många bokslutstransaktioner samtidigt, vilket också skapar problem om dessa multipliceras med annat än 1.