Avsnitt 5. Varför gör man prognoser och hur?

Ofta har man tid att förbereda sig innan man måste handla (framförhållning). Då kan man göra en **prognos** över läget då handlingen skall utföras och på effekten av ens handlande.

Kontrollerbart Plan | Handling | Prognos Okontrollerbart

Har man ingen framförhållning behövs ingen prognos. Då har man kanske en allmän handlingsmodell per automatik (eller panik!). Exempel: olycksfallsförsäkring. Prognos på jordbävning?

Varför hålls en kurs om prognoser i statistikprogrammet? Jo, för statistiken handlar om **inferens**. Vi har ett litet antal observationer och vi önskar dra slutsatser utanför samplet, dvs. om populationen. Ligger observationerna i tiden kan de på vissa villkor (stationaritet) betraktas som ett sampel ur en oändlig mängd möjliga historiska förlopp. Med hjälp av sampelestimat gör vi sannolikhetsuttalanden om tidsseriens förlopp utanför samplet, dvs. i framtiden.

Vi har blivit mycket bättre på att göra astronomiska och fysikaliska prognoser. T.ex. när asteroider passerar jorden, när en kastad sten träffar marken, osv. T o m. väderleksprognoserna har blivit träffsäkrare.

1. Identifiera prognosbehov (kort, lång, noggrannhet osv.)
2. Välj lämplig metod.
3. Anslåresurser.

Noggrannheten kan förväntas vara omvänt proportionellt mot **prognoshorisonten**.

Om det egna handlandet har stark inverkan på det som prognostiseras

är avsikten just att prognosen *inte* skall slå in.

Oftast gäller det att inte bara prognostisera en enda framtida punkt för en variabel utan att måa en bild där **många variabler bildar en helhet**, exempelvis en makroekonomisk prognos som utgår från ett slutet beräkningssystem: nationalbokföringen, dvs. Nationalräkenskaperna; inte bara BNP utan också dess komponenter.

Motsatsen, en ARIMA-modell är **univariat** och ser tidsserien som ett fenomen i en i övrigt tom rymd.

Kvantitativa – **kvalitativa** prognoser.

Mycket data kvantitativa. Lite data kvalitativa.

Exempel: Sveriges ekonomi, mycket data, numeriska prognoser.

Industrienkäten (Barometern) är snabbdata: går företaget *bättre – lika bra - sämre* .

Numeriska prognoser kan vara tidsserie-extrapolerade eller multivariata modeller baserade på ekonomisk teori.

Observera dock i boken:

BNP = f (penning- och finanspolitik, utländsk efterfrågan, löneuppgörelser,…, fel)

kan *inte* estimeras som en enda ekvation eftersom det finns feed-back från BNP på vissa av högerhands-variablerna. Man behöver ett simultant system av ekvationer.

Ekvationer behöver alltid något på höger sida. Hur vet man de variablernas framtida värden? Gör man prognoser på dem, varför inte då direkt på vänsterhands variabeln?

Kvalitativa prognoser har vanligen mindre användbarhet. Hur mäter man träff-säkerhet? Mindre transparenta än en numerisk modellprognos.

Sex steg i prognosarbetet (fem i boken):

1. Vad skall prognostiseras och hur skall arbetet bedrivas? T ex. produktionsprognoser, efterfrågeprognoser, prisprognoser…Hur skall man kunna studera fenomenet och insamla data och var skall analysen bedrivas?
2. Samla in både hårda data och erfarenhet hos de som jobbar i verksamheten.
3. Beskrivande analys. Försök förstå hur processen funkar.
4. Pröva olika modeller. Skall det vara något mycket enkelt eller kanske något som grundar sig på maximum likelihood modeller (ARIMA; VAR. Kalman osv.).
5. Utvärdera modellen genom att testa dess beteende dels under estimerings-perioden, dels utanför densamma på data som sparats för detta ändamål.
6. Se till att den bästa modellen införs som ett permanent prognosverktyg. Viktigt med uppföljning (se kap 5 i boken). Ett verktyg som alltid ligger på hyllan är inte nyttigt.

En prognos som helt utnyttjar all information om en variabels framtida förlopp kallas **rationell**.

Avsnitt 6. Verktyg för prognoser, träffsäkerhet, fördelnings-prognoser

När man fått ihop data (nu talar vi bara om kvantitativa prognoser) skall man alltid rita diagram. Man skall förstås ha observationer bevarade på ett vettigt sätt i det data-set man skall använda för sin analys, men bäst får man en uppfattning om data genom diagram.

[Figure 2.2]

Man ser då om det finns en trend, om data verkar vara konjunkturberoende och/eller ha en säsongvariation (bara om observationsintervallet är kortare än ett år, kvartal, månad, vecka osv.).

I Figure 2.2 ser man säsongvariation, vilket man kan vänta sig av ölförsäljning. Man tycker sig också ana en svagt sjunkande trend, men serien är lite för kort för att man skall kunna skilja så klart mellan trend och konjunktur. Men redan en dekomponering i:

$$y = TREND\&CYKEL \times SÄSONG \times SLUMPMÄSSIG$$

kan hjälpa en att göra en bättre prognos än en prognos som inte ser att en sådan dekomponering låter sig göras. Mer om det senare.

Fördelningar och statistikor – koncentrerad infomration

Ett annat sätt att studera data är att se på observationernas fördelning i form av **histogram**.

[Histogram över revideringar]

Ett förkortat sätt att karakterisera fördelningar är genom **statistikor**:

Lokaliseringsmått, dvs. en siffra som säger var på skalan fördelningen finns: aritmetiskt medelvärde, median, mod (typvärde) osv.

Spridningsmått visar hur utbredd fördelningen är: standardavvikelse, interkvartilavståndet, variationsbredden, medelkvadratfelet osv.

Har man fler än en variabel vill man kanske se på **korrelationen** mellan dem. Med bara en variabel är dess motsvarighet **autokorrelationen** för olika lags, som tillsammans bildar dessa autokorrelationsfunktionen, som man behöver för specificering av ARIMA-modeller.

När man väl hittat ett sätt att prognostisera vill man undersöka hur träffsäkra prognoserna är, dvs. man analyserar **prognosfelen** e_t. Detta beräknas som

$$e_t = Y_t - F_t,$$

Där Y_t är utfall och F_t är prognos, allt vi tidpunkt t. För att tydliggöra när prognosen för tidpunkt t är gjord brukar man skriva $F_{t|\,t\text{-}1,}$ som alltså är gjord vid t – 1. Detta kallas för en **ett-stegsprognos** (one-step-ahead forecast). Det är detta prognosfel som minimeras i ARIMA-modeller. Men med en ARIMA-modell och med många andra ansatser kan man också hoppa flera steg framåt rekursivt så att man använder modellen först med en riktig observation som utgångspunkt och därefter med start från ett-stegs-, tvåstegs- osv. prognosen. Allmänt kan man skriva $F_{\,t\,|\,t\,-\,k}$, för en prognos gjord för k perioder sedan eller $F_{\,t\,+\,k\,|\,t}$, om man vill poängtera att man utgår från dagens period. Detta kallas en **flerstegsprognos** (multi-step-ahead forecast). Observera att om man har en adekvat modell är ett-stegsprognoserna *okorrelerade*, medan flerstegsprognoserna automatiskt blir *(auto-)korrelerade*.

Mått på träffsäkerhet

Enklaste måttet, medelfelet: $ME = \frac{1}{n}\sum_{t=1}^{n} e_t$

Om prognosen träffar både över och under utfallet kunde man tänka sig att över- och underskattningar tar ut varandra och att *ME* då är noll. Är det inte det kanske det är fråga om **bias**, dvs. att prognosen inte är förväntningsriktig, utan antingen i genomsnitt alltför pessimistisk eller optimistisk. Däremot säger ju måttet ingenting om hur stora felen kan vara.

Om felens fördelning är någotsånär symmetrisk är medelfelet och **medelkvadratfelet**

$$MSE = \frac{1}{n}\sum_{t=1}^{n} e_t^2$$

bra karakteristikor för denna fördelning. För att överföra denna karakteristika till samma skala som observationerna brukar man ta **kvadratroten** av *MSE* och beteckna den som *RMSE*. Ibland lägger man till ett *"F"* för forecast och då blir det *RMSFE.* Detta är det vanligaste spridningsmåttet. Men är fördelningen sned är det bättre att ange medianen

Med = Mittersta observationen om n är udda, annars medeltalet av de två mittersta.

Det motsvarande spridningsmåttet är det absoluta medelfelet

$$MAE = \frac{1}{n}\sum_{t=1}^{n} |e_t| \ .$$

Är felfördelningen symmetrisk blir $ME = Med$ och $RMSE \cong 1.25\, MAE$.

En nackdel med *RMSE* och *MAE* är att de är beroende av den prognosticerade tidsseriens varians. Man kan alltså inte göra en rättvis jämförelse av prognoser mellan variabler med olika varians med hjälp av dessa karakteristikor. Det är klart att det är svårare att prognosticera en variabel med stor spridning än en med liten. Är spridningen noll är det ju urenkelt!

Därför brukar man normera *RMSE* med variabelns estimerade standardavvikelse s. Vill man normera *MAE* bör man använda motsvarande karakteristika för variabeln, dvs. den absoluta medelavvikelsen från medianen:

$$MAD = \frac{1}{n}\sum_{t=1}^{n}\left|e_t - Med\right| .$$

Ibland räknar man även dessa mått i procentenheter. Då blir de också bättre jämförbara mellan olika variabler. Vanligast är det **procentuella prognosfelet**

$$PE_t = \left(\frac{Y_t - F_t}{Y_t}\right)\times 100 ,$$

medelprognosfelet

$$MPE = \frac{1}{n}\sum_{t=1}^{n} PE_t$$

och **absoluta procentuell medelfelet**

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|PE_t\right| .$$

Procentuella fel är lättare att förstå än RMSE, som ju alltid måste relateras till den skala tidsserien ligger i.

Att jämföra prognoser

Hur vet man om en prognos är bättre än en annan? Eller om en prognos överhuvudtaget är användbar? Vi tar ett fall då dessa frågor sammanfaller. För att ha informationsvärde måste man kräva att prognosen är bättre än något mekaniskt som man kunde lära en apa att göra. Man definierar olika slag av **naiva prognoser** att jämföra med.

Naiv 1: $F_t = Y_{t-1},$

dvs. prognosen för *t* är helt enkelt lika med den sista observationen. Detta är faktiskt en optimal prognos i det fall att Y_t är en **slumpvandring** (random walk) och inget

samband med någon annan variabel är känd. Vi vet att de flesta ekonomiska tidsserier innehåller autokorrelation och denna information utnyttjas alltså inte i *Naiv* 1. Dessutom finns det ofta annan information som kunde utnyttjas.

Har serien stark säsongvariation kan man sätta prognosen lika med samma periods värde ett år innan. I fallet kvartalsdata får man

Naiv 2: $F_t = Y_{t-4}$.

Dessa nämns i boken. Jag lägger till ytterligare en,

Naiv 3 : $F_t = \frac{1}{t-1}\sum_{t=1}^{t-1} Y_{t-1}$,

Dvs. medelvärdet för observationerna man just då har.
Man kan jämföra *RMSE* för den prognos man gjort med *RMSE* för någon av de naiva metoderna. Detta kallas **Theils** U **- statistika**. Det är alltså inte fråga om ett test, bara en deskriptiv jämförelse. För *Naiv* 1 får man

$$U = \sqrt{\frac{\sum_{t=2}^{n}(F_t - Y_t)^2}{\sum_{t=2}^{n}(Y_t - Y_{t-1})^2}} = RMSE(\text{modell})/RMSE(\text{föreg. värde}) .$$

$U < 1$ är en indikation på att prognosen kan vara bättre än en naiv ansats och vice versa. Konfidensen är i detta fall okänd. Vill man göra en hypotes av dessa indikationer blir den

H_0: $U = 1$

H_1: $U < 1$.

Ett test man kan använda sig av för att pröva H_0 kallas Granger-Newbold testet. Kalla felen för metod 1 e_{1t} och felen för metod 2, e_{2t}. Testa så korrelations-koefficienten mellan summan för och skillnaden mellan dessa, alltså

$$r = \frac{\sum (e_{1t} - e_{2t})(e_{1t} + e_{2t})}{\sqrt{\sum (e_{1t} - e_{2t})^2 \sum (e_{1t} + e_{2t})^2}}$$

och $$t = \frac{r\sqrt{n-2}}{1-r^2}, df = n-2.$$

Vad är nu tanken bakom detta test? Jo, egentligen testar vi ju

$$H_0: Ee_{1t}^2 = Ee_{2t}^2,$$

för då är ju prognoserna i genomsnitt lika träffsäkra. Men täljaren i

$$r = (e_{1t} - e_{2t})(e_{1t} + e_{2t}) = e_{1t}^2 - e_{2t}^2,$$

så att $Er = Ee_{1t}^2 = Ee_{2t}^2$. Korrelationskoefficienten mäter alltså skillnaden i träffsäkerhet.

Prognosfelens autokorrelation

Vi måste tillåta prognosfel, för en prognos är exakt rätt bara av en ren slump. Däremot måste vi kräva att modellen uppfyller de villkor vi ställde upp för en modell,

$$y = M + a$$

där alltså felen a är helt slumpmässiga, där all systematik finns med i M och där dessa två är sinsemellan oberoende. Finns det autokorrelation i $\hat{a}$ är detta en systematik som gör hemma i som används för att göra prognoser och som blir träffsäkrare med denna information inbyggd.

Uppfriskning av minnet, autokorrelationen r_h för lag h är

$$r_h = Cov(y_t, y_{t-h})/Var(y_t).$$

Diskutera varför formeln ser ut som den ser ut. För h = 1, 2, …, n p får man **autokorrelationsfunktionen** *(ACF).* Figure 2-7 ser vi att det finns Information att hämta ur residualen på säsongsfrekvensen 12 månader.

[Figure 2-7]

ACF kan ju inte vara identiskt noll annat än av slumpen så en skattad funktion har alltid värden olika noll. En tumregel är att man bara behöver bry sig om ett fåtal lags, Men med beaktande av perioden i periodiska data och bar sådana som är större än $\pm 2/\sqrt{n}$, där n är antalet observationer.

Prognosintervall

Under förutsättning att prognoserna kan anses vara förväntningsriktiga och felen normalfördelade kan man lägga konfidensintervall kring punktprognosen. Som vänligt väljer man värdet z för en konfidens man bestämt sig för ut normalfördelningen, för 5% är det 1.96. Intervallet blir då

$$F_{t+1} = \pm z \cdot RMSE.$$

Observera att detta bara gäller för en ett-stegs-prognos. Det är svårare att göra flerstegsprognoser. Då växer intervallet ju mer man avlägsnar sig från dagsläget. Man kan också välja andra värden från N-fördelningen. Som minnesregel kommer 2/3 av prognoserna att ligga i intervallet $\pm RMSE$ och ung. 95% i $(-2 \cdot RMSE, 2 \cdot RMSE)$.

Vad man egentligen försöker göra då man beräknar konfidensintervall är att försöka ge en bild av den fördelning man egentligen projicerar in i framtiden. Man talar då om en **fördelningsprognos**. Det har också börjat förekomma att man publicerar fördelningsprognoser, isynnerhet om den inte är symmetrisk. Riksbanken har en sådan inflationsprognos (bild).

[Inflationsprognos]

Det kan också hända att beslutsfattarens förlustfunktion inte ser ut som i Figure 2-8 (kvadratisk), utan det kan kosta mer att göra fel t ex. till vänster, så att undervärdering kostar mer än övervärdering. Då gäller inte RMSE och N-fördelning.

Dt är också viktigt att testa ifall prognosen kan anses vara förväntningsriktig och N-fördelningsantagandet innan an använder intervallskattningar.

[Figure 2-8]

Transformationer

Oftast är ekonomiska tidsserier växande i tiden (produktion, priser, arbetskraft osv.). Då kommer trenden att böja sig uppåt, bli exponentiellt växande. Variansen ökar också med nivån (Figure 2-10 och 2-11).

[Figure 2-10 och 2-11]

Det vanliga är då att man logaritmerar tidsserien. Man får då direkt estimat på elasticiteter i regressionsekvationer med ekonomiska variabler. För små förändringar ligger differensen av en log. Serie nära procentuell förändring, som är ett väl inarbetat mått. Men viktigast är att man kan få serien **homoskedastisk**, dvs. variansen är konstant över tiden. Det är mycket svårare att arbeta med **heteroskedastiska** serier.

En generellare transformation introducerades av George Box och David Cox. Den kallas Box-Cox transformationen:

$$y = x^{\ddot{e}} / \ddot{e} .$$

När $\ddot{e} \quad 1$ blir det ingen transformation och när $\ddot{e} \quad 0$ får man log.

Prognoser i statistisk tid och i real tid och osäker statistik

Vi vet att det tar tid att göra statistik över ekonomins utveckling. Därför är även den färskaste statistiken på månader 1 ½ månad gammal och kvartalsstatistiken ett kvartal gammal. Då försöker komma närmare nuet med **ledande indikatorer** (leading indicators) och sk. **Flash-estimat**, även kallade ”nowcasts”. Dessa behöver inte vara prognoser i realtid utan de skall komma snabbare än statistiken. Det räcker. Prognoser i realtid försöker se längre än nuläget.

När så statistiken kommer är det i form av preliminära värden, som kan vara mycket osäkra. Här kommer några exempel från Öller & Hansson (2002).

[Histogram, konjunkturbild, säsong, internationellt]

Kontentan av detta blir att det är svårt att kontrollera hur bra en prognos är i förhållande till utfallet. Den bild man har är att en prognos försöker sikta på ett stillastående mål. Liksom i prickskytte kan man gå och mäta avståndet mellan prognos och mål. Detta är vårt e_t . Men vilket utfall skall man välja? Det preliminära har vi sett att revideras mycket under två års tid. Men skall vi ha som mål den slutgiltiga siffran som är närmast inaktuell när den äntligen kommer två år senare? Hur skall man ens veta om en reviderad siffra faktiskt ligger närmare ett okänt ”riktigt värde”. Det finns gott om belägg för att revideringar syns ha gått åt fel håll. Det man gör när man räknar prognosfel är egentligen att man jämför två estimat.