

SKÖRDARNA OCH DERAS INVERKAN PÅ 1800-TALETS STÖLDBROTT

Ett genmäle om regressionsanalys och regression i analys.

I NTfK 1985: 2 presenterade Johannes Knutsson bl a en regressionsanalys ang sambandet mellan lagföringar för stöld och skördeutfallet under åren 1860-1889. Han drar slutsatsen, att man med hjälp av en regressionsmodell kan påvisa ett samband mellan stöld och skördeutfall och skriver sammanfattande:

1. »Under perioden 1860-1889, med en kraftig uppgång av brottsligheten åren 1867-1870 och den sedan nedåtgående trenden, uppvisas således en samvariation med skördeutfallet. De övriga smärre variationerna tas väl om hand av skördevariabeln /.../« (s 106).

2. »Variationerna i 1800-talets stöldbrottslighet kan återföras till grundläggande ekonomiska förhållanden i befolkningen. /.../ Enligt denna studie skulle både de kortsiktiga variationerna och nivån av 1800-talets stöldbrottslighet kunna återföras till en faktor/skördeutfallet, HvH/ som påverkade människors motivation till att begå brott« (s 109).

Knutsson kan ha rätt med dessa påståenden, men de har inget stöd i hans data. När det gäller påstående 1 gäller bristerna den statistiska analysen. När det gäller påstående 2, vilket i sak rör frågan om stöldbrottslighetens *nivå* under senare delen av 1800-talet, är det istället fråga om att Knutsson tillskriver regressionsansatsen förklaringsmöjligheter som den inte har.

Påstående 1: De övriga smärre variationerna tas väl om hand av skördevariabeln.

Knutsson redovisar tyvärr inga rådata – en vanlig brist i framställningar av detta slag – vilket försvårar granskningen av resultaten. Jag var tvungen att ta ett diagram i ett tidigare papper i samma ämne (Knutsson, 1978) till hjälp för att kunna få fram värdena på skördevariabeln.¹⁾

Det enklaste sättet för att kontrollera om stölderna samvarierar med skördeutfallet är att rita *upp- och nedgångarna* i stöld- och skördevariabeln som »1« och »0« på ett papper.²⁾ Då ser man (Fig 1) att de enligt hypotesen förväntade omaka kombinationerna mellan stöld och skörd (1/0 resp 0/1) uppkommer i högst 19 av 28 möjliga försök (två fall i slutet av 1880-talet går dock egentligen inte att avgöra).

Fig 1: Upp- och nedgångar i stöld- och skördevariabeln, 1862-1889.

Skörd	0 1 0 1 0 0 0 1	1 0 1 0 1 0 1 0 1 1	0 1 0 1 0 1 0 - 1 -
Stöld	1 0 0 0 1 1 1 0	0 0 0 0 0 1 0 0 1 0	1 1 0 0 1 0 0 0 1 1
	1860-talet	1870-talet	1880-talet

1) För enkelhetens skull inskränker jag mitt fortsatta resonemang på endast männens stöldlagföringar och skördeutfallet (med 1-års lag). Rekonstruktionen av skördevariabeln för åren 1860-1888 bör kunna anses som lyckad, vilket följande jämförelse visar:

Knutssons skattning: $216,99 - 2,58 \text{ skörd}(t-1) + 66,08 \text{ missväxt}$.

Min skattning: $216,99 - 2,58 \text{ skörd}(t-1) + 66,08 \text{ missväxt}$.

2) För att undvika missförstånd: denna mycket grova, men enkla metod är endast tänkt som en första översiktlig kontroll för i vilken riktning sambandet går; den ska *inte* ersätta eventuella parametriska modellskattningar.

Resultatet bestäms emellertid i hög grad av utfallet under 1860-talet. Då uppkommer 7 omaka kombinationer av 8 möjliga; under 1870-talet däremot endast 6 av 10 och under 1880-talet i bästa fall 6 av 10 (två fall egentligen oavgjorda). Dylika resultat för 1870- och 1880-talen skulle man också kunnat få genom att singla slant.

Redan under 1870-talet tycks alltså ett statistiskt säkerställt samband mellan skördarnas och stöldernas *årliga* upp- och nedgångar ha försvunnit.

Hur kan då Knutsson komma fram till resultat som motsäger en enkel inspektion av data? Förklaringen ligger i det att Knutsson valde en förhållandevis anspråksfull ansats, som kräver varsam hantering och tolkning för att kunna ge rätt resultat. ³⁾ Jag tänker kritisera Knutsson i tre hänseenden:

- a) ang tolkningen av innebörden av det s k R^2 -värdet;
- b) ang hanteringen av dummyvariabeln;
- c) ang inkluderingen av trenden.

Ad a: R^2 -värdet

Att Knutsson dragit slutsatsen, att stöldlagföringarnas variationer väl tas om hand av skördevariabeln, hänger uppenbarligen samman med att han i datautskriften förlitade sig på modellens »förklaringsvärde«, dvs R^2 -värdet (se tablån på sid 106 i Knutssons artikel). Att det ligger på över 80 procent beror framför allt på att åren 1867-70 svarar för 73 procent av stöldvariabelns totala variation (= summan av de kvadrerade avvikelserna från medelvärdet). Dessa år har Knutsson dock speciellt pekat ut (»specificerat«) med hjälp av en dummyvariabel, varför det inte är konstigt, att den av modellen redovisade variationen blir så hög. Detta säger dock ingenting om, i vilken mån de årliga upp- och nedgångarna är »väl omhändertagna«, vilket framkommer tydligt när man jämför det faktiska utfallet med det av Knutssons modell predicerade (Fig 2). Fr o m 1873 förutsågs bara cirka hälften av de årliga upp- och nedgångarna rätt.

Fig 2: Stöld, 1862-1889. Faktiskt och predicerat utfall enl Knutssons modell.

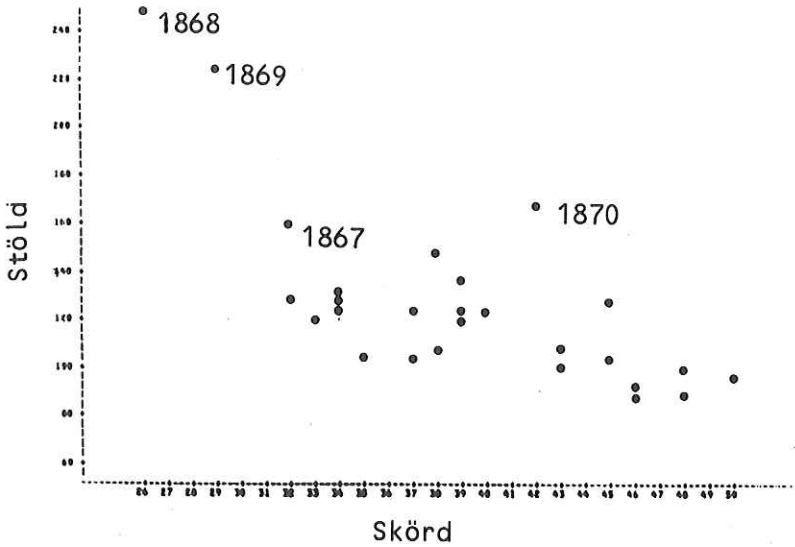
Fakt. utfall	1 0 0 0 1 1 1 0	0 0 0 0 0 1 0 0 1 0	1 1 0 0 1 0 0 0 1 1
Pred. utfall	1 0 1 0 1 1 1 0	0 0 0 1 0 1 0 1 0 0	1 0 1 0 1 0 1 - 0 -
	1860-talet	1870-talet	1880-talet

Ad b: Dummy-variabeln

För att värdena under 1860-talets svältår »inte ska förrycka resultaten« (dvs skattningen av skördecoefficienten; s 110), specificerade Knutsson *en* dummyvariabel för åren 1867-1870. I princip är det rätt att använda sig av denna teknik (se t ex Wonnacott-Wonnacott, 1977: 385 ff), men granskar man Fig 3, så ser man att dessa år i sig är mycket olika. Dels ligger stölderna åren 1868 och 1869 på över 200, medan värdena är kring 160 åren 1867 och 1870 (medelvärdet för hela perioden är 125).

3) Jfr Draper-Smith (1966:242): »... no scientist should be persuaded to abandon his scientific insight and principles in favor of some computerized statistical screening procedure. The use of multiple regression techniques is a powerful tool only if it is applied with intelligence and caution.«

Fig 3: Stöld vs skörd, 1861-1889



Dels är skördeåret 1869 (= lagföringsår 1870) inte något missväxtår – tvärtom tillhör det de åren med *goda* skörderesultat. Det är således inte självklart att behandla åren 1867-1870 lika. Utifrån Fig 3 skulle man exempelvis lika gärna kunna hävda, att år 1866 (= lagföringsår 1867) *inte* bör särskilt specificeras samt att skördeår 1869 (= lagföringsår 1870) bör ges en *egen* dummy-variabel.

Skillnaden mellan dessa och Knutssons antaganden för skattningen av skördecoefficienten blir betydelsefulla när sedan också trenden i materialet tas i beaktande.

Ad c: Trenden

Knutsson har inte korrigerat för trenden i materialet, fastän han själv skriver: »Under perioden 1860-1889, med en kraftig uppgång av brottsligheten åren 1867-1870 och *den sedan nedåt-gående trenden* /.../« (s 106; min understrykning). Det finns två skäl varför man bör korrigerar för trender. Dels förutsätter Knutssons val av statistisk modell trendfrihet; dels är trender utan några tydliga avbrott intetsägande i tolkningssammanhang (mycket åskådligt i Yeomans, 1968:232 ff).

För trendkorrigeringar finns det åtminstone två standardmetoder. Antingen lägger man in en trendvariabel av dummy-typ (Draper-Smith, 1966: 137 ff) eller så bildar man s k första differenser (ofta använt inom tidsserieanalys; se t ex Pankratz, 1983:24 ff).

Räknar man på det första sättet, skattas skördecoefficienten i Knutssons modell till $-2,19$ ($t = -2,58$) och i den ovan beskrivna alternativa modellen till $-0,69$ ($t = -2,11$)⁴. Väljer man i stället de första differenserna, antar skördecoefficienten värdet $-0,64$ ($t = -2,22$)⁵. Detta värde är ganska lika skattningen av skördecoefficienten i den alternativa modellen, vilket kan tas som intäkt för att den inte är helt orimligt specificerad.

Oavsett på vilket sätt man försöker specificera dummyvariablerna, lyckas dock ingen uppsättning på ett tillfredställande sätt förutsäga stödlagföringarnas årliga *upp- och nedgångar* under 1870- och 1880-talen.⁶ Detta bör tolkas så, att antingen utfaller stödlagföringarnas årliga variationer slumpmässigt (vilket jag anser som det troligaste) eller så fattas det en eller flera variabler i modellen. Frågan kan undersökas närmare, men det ska inte göras här. I vilket fall som helst är det *inte* skördarna som i första hand bestämmer stödlagföringarnas årliga variationer under 1870- och 1880-talen.

Vad blir då sammanfattningsvis slutsatsen av Knutssons statistiska undersökning ang åren 1861-1889?

* Det var missväxt i slutet av 1860-talet med en kraftig ökning av stödlagföringar till följd. Detta visste vi förut (t ex Hellstenius, 1871; Knutsson, 1978).

* Skördarna blev bättre under senare delen av 1800-talet och stölderna minskade. Också detta visste vi förut (t ex Widell, 1904; Knutsson, 1978) och det kan även direkt avläsas ur statistiken. Några komplicerade beräkningar är obehövliga och tycks i det här fallet fördunkla mer än de klargör. Trender utan några tydliga avbrott är dessutom intetsägande i tolkningssammanhang. Utvecklingen kan lika gärna ha bestämts av exempelvis reallönernas utveckling eller andra faktorer. Några av dem nämnde också Knutsson.

* Redan under 1870-talet finns inget statistiskt säkerställt samband kvar mellan skördarnas och stöldernas *årliga* upp- och nedgångar, när man korrigerat för trend. Det är missvisande att behandla 1860-talet och 1870- och 1880-talen som en enhetlig analysperiod. Att förlägga »tidpunkten för den direkt nödbaserade stöldbrottslighetens försvinnande« till omkring 1890 har inte täckning i data. Helt bortsett ifrån att den låga stöldnivån i slutet av 1880-talet, som av ögat lätt uppfattas som en vändpunkt, *kan* ha mycket speciella förklaringar: små ungdomskullar som en följd av missväxten 20 år tidigare (!) och hög utvandring. Förmodligen är det rimligare att tolka slutet av 1880-talet som en *tillfällig* avvikelse från en i övrigt mer eller mindre stabil stöldnivå under perioden 1871-1939.

4) Efter korrigerig för autokorrelation avs lag 1 och 2 (metod: Yule-Walker). För både Knutssons och den alternativa modellen gäller: Trend = 1,2, ..., 19 för åren 1871, 1872, ..., 1889; 0 annars. Tidpunkten bestämdes utifrån Knutssons formulering »och den sedan nedåtgående trenden« (s 106). På grund av att materialet endast omfattar få (29) observationer är dessa och de följande skattningarna inte speciellt robusta.

5) Efter en motsvarande dummy-korrigerig för åren 1868 resp 1870 och för autokorrelation avs lag 1 och 2 (interceptet satt till 0).

6) På motsvarande sätt som ovan (inklusive tidsvariabeln) skattas skördecoefficienten för perioden 1871-89 till $-0,53$ ($t = -1,22$) efter korrigerig för autokorrelation avs lag 1 och till $-0,46$ ($t = -1,20$) enl första-differens-metoden. Detta påminner om resultatet från 0/1-analysen ovan: också den gav en negativ, men inte signifikant skördecoefficient för 1870- och 1880-talen.

Påstående 2: Det är skördeutfallet som bestämmer stöldernas *nivå* under 1800-talet.

Knutsson har förmodligen rätt, när han skriver att skördevariabeln under 1800-talet genomgått ett »kvalitativt skifte«. ⁷⁾ Han borde emellertid ha lagt till, att *stölderna* troligen inte gjort det, eftersom undersökningen gällde i vilken mån *stölderna* under 1800-talet kan anses vara »nödbaserade« (s 102).

Även om variationen mellan stöld och skörd efter missväxtåren i slutet av 1860-talet hade varit fullständig, så skulle detta ändå säga ganska lite om stöldernas eventuella nödbasering, eftersom stölderna under perioden 1871-1889 årligen inte varierat med mer än genomsnittligt +/- 7 procent (max +16/- 21 procent). Analysen av tidsseriedata kan endast ge information om variationen kring medelvärdet (»stöldernas årliga variationer«), men den kan inte säga någonting om medelvärdets nivå (»stöldernas övriga orsaker«). Rent allmänt »förklarar« regressionsanalyser inte nivån hos den analyserade variabeln; den »förklarar« däremot skillnaden mellan ett godtyckligt par av observationer (Lütjohann, 1984:19).

Oavsett hur man ställer sig till skördeutfallets betydelse för stöldlagföringarnas *årliga* variationer antyder tvärsnittsuppgifter i rättsstatistiken, att de straffade för stöld (exkl snatteri) även långt efter 1890 rekryterades från de befolkningsskikt som betecknades leva under »knappa« och »utfattiga« förmögensvillkor. År 1912 rapporteras exempelvis fortfarande 80 procent komma från »knappa« eller »utfattiga« förmögensvillkor (BSOS 1912: Tab 6).

Jag frågar mig, vad som kan ha varit Knutssons avsikt med att publicera sin artikel – speciellt mot bakgrund av att ett tidigare – omsorgsfullt resonerande och läsvärt – papper (1978) i samma ämne redan har blivit uppmärksammat och citerat även utanför kriminologernas trånga krets (se t ex Petersson, 1983). Inte heller kan man påstå att själva ämnet skulle vara en vit fläck på forskningskartan. En ompublicering kunde kanske ha blivit meningsfull, om Knutsson valt att ge oss en systematisk genomgång av svenska och internationella studier i frågan (se t ex Council of Europe, 1984). Han kunde också ha tagit tag i den – även för dagens kriminologi – mycket viktiga perioden i slutet av 1840-talet fram till slutet av 1850-talet för att öka våra kunskaper. Här minskar ju lagföringarna för stöld mycket dramatiskt (med mer än 50 procent). Är det (bara) skördarna som har betydelse? Vilken roll spelar exempelvis förbudet av husbehovsbränningen under 1850-talet? Eller fängelsestraffets införande som regelstraff för stöld vid ungefär samma tidpunkt? Författaren nämner inte ens dessa händelser.

Referenser

BSOS (1912), Bidrag till Sveriges Officiella Statistik. B) Rättsväsendet. Ny följd. LV: 2. / ... / År 1912. Stockholm 1913.

Council of Europe (1984), Historical Research on Crime and Criminal Justice. Sixth Criminological Colloquium. Strasbourg 1984.

N. Draper – H. Smith (1966), Applied Regression Analysis. New York: John Wiley & Sons, 1966.

J. Hellstenius (1871), »Skördarna i Sverige och deras verkningar«, Statistisk tidskrift 1871, ss 77-119.

7) Fastän dåliga skördar kan fortfarande haft en viss betydelse för stöldbrottsligheten så sent som under krisåren i slutet av 1:a världskriget.

- J. Knutsson* (1978), Nöd som förklaring till 1800-talets tillgreppsbrottlighet (II). Stockholms universitet. Allmän kriminologi (utkast). Stockholm juni 1978 (stencil).
- H. Lütjohann* (1984), Regressionsanalys för f d statistikstuderande. Promemorior från P/STM. Nr 12. 1984-02-01. Statistiska centralbyrån.
- A. Pankratz* (1983), Forecasting With Univariate Box-Jenkins Models. Concepts and Cases. New York: John Wiley & Sons, 1983.
- B. Petersson* (1983), »Den farliga underklassen«. Studier i fattigdom och brottlighet i 1800-talets Sverige. Acta Universitatis Umensis 53. Umeå: Almqvist & Wiksell, 1983.
- L. Widell* (1904), »Bidrag till Svensk Kriminalstatistik«, Statistisk tidskrift 1904, ss 139-285.
- T. H. Wonnacott - R. J. Wonnacott* (1977), Introductory Statistics. Third Edition. New York: John Wiley & Sons, 1977.
- K. A. Yeomans* (1968), Introducing statistics. Statistics for the social scientist: 1. Harmondsworth: Penguin Books, 1968.

Data

År	Stöld	Skörd	År	Stöld	Skörd	År	Stöld	Skörd
			1870	166	42	1880	119	39
1861	125	34	1871	147	38	1881	128	45
1862	128	32	1872	137	39	1882	124	37
1863	124	39	1873	125	34	1883	98	48
1864	119	33	1874	124	40	1884	100	43
1865	105	37	1875	126	34	1885	97	50
1866	133	34	1876	109	43	1886	91	46
1867	159	32	1877	104	35	1887	86	46
1968	250	26	1878	108	38	1888	87	48
1869	225	29	1879	103	45	1889	89	48

Angående definitionerna, se ovan not 1).

Adresse: *Hanns v Hofer*

Statistiska centralbyrån, S-115 81 Stockholm