

Rapport 2003:4

En snabb BNP-indikator för det privata näringslivet på Åland

The logo for ÅSUB, featuring the text 'ÅSUB' in a bold, white, sans-serif font. The letter 'Å' has a small circle above it. The text is set against a dark blue background with a white border.

ÅSUB

De senaste rapporterna från ÅSUB

- 2002:3 Konjunkturläget - våren 2002
- 2002:4 Konjunkturbedömning för den kommunala sektorn våren 2002
- 2002:5 Den åländska näringsrätten
- 2002:6 Efterhandsutvärdering Ålands mål 5b-program 1995-1999
- 2002:7 Konjunkturbedömning för den kommunala sektorn - hösten 2002
- 2002:8 Konjunkturläget hösten 2002
- 2002:9 Bussiga stan
- 2003:1 Konjunkturläget våren 2003
- 2003:2 Det åländska jordbrukets framtida utvecklingsförutsättningar
- 2003:3 Konjunkturbedömning för den kommunala sektorn - våren 2003

Pris 10,00 €

ISSN 1455-1977

Förord

Ålands statistik- och utredningsbyrå (ÅSUB) har under en längre period aktivt arbetat med att utveckla och förbättra dataunderlaget för sina konjunkturanalyser. Arbetet med de årligen återkommande konjunkturöversikterna har också blivit allt mer omfattande. Sedan ett par år tillbaka publicerar ÅSUB varje år fyra konjunkturrapporter, två under våren och två under hösten. Två av dessa gäller konjunkturläget inom den åländska ekonomin som helhet och två är speciellt inriktade på den kommunala sektorns behov. De två förstnämnda vår- och höstrapporterna tas fram på uppdrag av finansavdelningen vid Ålands landskapsstyrelse, medan de sistnämnda tas fram på uppdrag av Ålands kommunförbund.

Konjunkturarbetets växande omfattning i kombination med ÅSUBs begränsade (personal-) resurser har gjort det nödvändigt att försöka hitta dataunderlag och prognosmodeller som så långt som möjligt begränsar arbetsinsatserna samtidigt som bästa möjliga resultat ändå kan säkras. Som ett led i denna utveckling har användningen av olika typer av företagsbaserade registermaterial utökats och förfinats, samtidigt som de så omfattande och arbetskrävande företagsenkäterna från och med innevarande år begränsats till att göras endast en gång per år - mot tidigare två gånger per år. En förutsättning för detta är dock en effektivare och mer precis användning av andra typer av information och data, och då inte minst ovan nämnda slag av registerdata.

För att stöda arbetet med att förbättra utnyttjandet av befintliga registerdata - och därmed öka precisionen i ÅSUB's prognoser över den åländska ekonomins utveckling - knöts kontakter med den statistiska institutionen vid Linköpings universitet. Dessa resulterade i att statistikstuderande *Kalle Wahlin* bestämde sig för att ägna sin magisteruppsats (D-uppsats) åt en analys av möjligheterna att på basen av data i det månatligen uppdaterade omsättningsskatterregistret utveckla en modell för framtagning av snabba BNP-indikatorer gällande Ålands privata näringsliv. Tanken var att man genom att använda "snabba data" om företagens omsättning, löneutbetalningar, moms och liknande i en metodiskt genomarbetad och empiriskt testad modell skulle få fram rimligt tillförlitliga prognoser över den privata sektorns bidrag till Ålands BNP-utveckling.

Föreliggande rapport är resultatet av detta arbete. Här redovisas och testas mot den faktiska ekonomiska utveckling en modell som möjliggör produktion av snabba BNP-indikatorer för den privata delen av den åländska ekonomin. Som framgår av rapporten genererar modellen användbara BNP-indikatorer för näringslivet som helhet men dock inte gällande enskilda branscher och näringsgrenar. Resultatet bekräftar också den åländska makroekonomins stora beroende av en handfull storföretag: en kraftigt reducerad modellkörning omfattande enbart de tio största företagen ger nämligen ungefär samma prognosutfall.

Den i rapporten presenterade modellen har i en reducerad version redan tillämpats i samband

med den senaste omgången konjunkturöversikter (våren 2003). Den av Wahlin framtagna och testade huvudmodellen kommer dock att tas i mera reguljär användning redan i samband med höstens konjunkturöversikter.

Rapporten är till sin karaktär betydligt mer teknisk än vad som är fallet i ÅSUB's mer ordinarie publikationer. Eftersom den därtill producerats i ett externt utbildningssammanhang, så vilar ansvaret för textens innehåll i detta fall huvudsakligen på författaren. Rapportens grundläggande dokumentation av de metodologiska förutsättningarna för denna del av kommande BNP-prognoser (vilka även bygger på andra typer av material), gör den dock till en väl motiverad produkt i ÅSUB's utredningsserie.

Som beställare och initiativtagare till rapporten är ÅSUB nöjd med resultatet och vill därför rikta ett tack för gott och resultatrikt samarbete, inte bara till Wahlin, utan även till av projektet berörda lärare och handledare vid Linköpings universitets statistiska institution. ÅSUB avser för sin del att även fortsättningsvis vidareutveckla den här typen av produktiva kontakter med studerande vid lämpliga akademiska institutioner.

En annan nyckelperson i detta projekt har varit statistiker/ekonom *Jouko Kinnunen* som varit den huvudansvarige inom ÅSUB för samarbetet med och handledningen av Wahlin. Kinnunen är också den som ansvarat för de preliminära modelltillämpningar som hittills gjorts i samband med ÅSUB's konjunkturanalyser.

Mariehamn i augusti 2003

Bjarne Lindström
Direktör

Abstract

The gross domestic product (GDP) on the Åland islands is calculated by employing the principle of value added. According to this principle, the contribution of a firm to the GDP is defined as the difference between the turnover and the purchases from other firms. Information about turnover and purchases from other firms is available from the register of annual accounts established by the Finnish Tax Administration. However, this register is made available with a time lag of at least one year, and there is a need for more rapid information about the economical situation. In addition to the register of annual accounts there is a monthly register of value-added tax compiled and updated by the Finnish Tax Administration (together with Statistics Finland). The aim of this project was to use the latter register to construct a rapid method to predict the annual changes in the contribution of the private business to the gross domestic product. The problem was solved by fitting a regression model to the data. More precisely, preliminary GDP values were computed by employing a mixed linear model with the turnover, the payments to the social security system and the value-added tax as explanatory variables. Data for 1996-1999 were used to fit the prediction model, and the predictions for 2000 and 2001 indicated a small rise in the value added for the private business.

Sammanfattning

Ålands bruttonationalprodukt (BNP) beräknas enligt förädlingsvärdesprincipen. Enligt denna princip beräknas ett företags bidrag till BNP som differensen mellan företagets omsättning och värdet av dess insatsförbrukning. ÅSUB erhåller uppgifter om omsättning och insatsförbrukning ur Finlands Skattestyrelses årliga bokslutsregister. Detta register är dock tillgängligt först efter en tidsfördröjning av minst ett år. Därför finns ett stort behov av snabbare information om det ekonomiska läget. Förutom bokslutsregistret sammanställer och uppdaterar Skattestyrelsen (tillsammans med Statistikcentralen i Finland) ett månatligt momsregister. Detta projektarbete syftar till att bestämma en metod för att prognostisera de årliga procentuella förändringarna av det privata näringslivets tillskott till BNP innan uppgifterna från bokslutsregistret är kända. Uppgiften löstes genom att en regressionsmodell av typen mixad linjär modell anpassades till data. I den slutgiltigt valda modellen ingick omsättning, socialavgiftsinbetalningar samt moms till godo från alla inköp som förklarande variabler. Data för år 1996 – 1999 användes för att anpassa modellen, och sedan användes momsregisteruppgifter för år 2000 och 2001 för att prognostisera förädlingsvärdet för dessa år. Prognosen pekade på svaga öknings av det privata näringslivets tillskott till BNP för både år 2000 och 2001.

Innehållsförteckning

1. Inledning	6
1.1 Om beställaren.....	6
1.2 Bakgrund	6
1.3 Syfte och problemformulering	7
1.4 Omfattning	7
1.5 Disposition	8
2. Datamaterial och preliminär bearbetning	9
2.1 Konstruktion.....	9
2.2 Beskrivning	10
2.2.1 Datatyp	10
2.2.2 Numerisk beskrivning.....	11
2.2.3 Fördelning.....	13
2.3 Bearbetning	14
2.4 Felkällor i datamaterialet.....	14
3. Metod.....	16
3.1 Modell	16
3.2 Skattning.....	20
3.3 Prognostisering.....	20
4. Utförande	22
4.1 Förklarande variabler	22
4.2 Modellval	25
4.3 Parameterskattning.....	28
4.4 Modellens validitet.....	30
4.5 Felkällor i modellen	31
5. Tillämpning.....	34
5.1 Branschvis	34
5.2 Hela det privata näringslivet	40
6. Reducerad metod.....	42
6.1 Prognos.....	42
6.2 Jämförelse.....	45
7. Slutdiskussion	46
Referenslista.....	48

1. Inledning

1.1 Om beställaren

Ålands statistik- och utredningsbyrå (ÅSUB) är en fristående organisation vars huvuduppgift är att verka som officiell statistikmyndighet på Åland samt att bedriva forsknings- och utredningsverksamhet. ÅSUB producerar statistik inom en rad samhällsområden. Tyngdpunkten ligger på utredningar inom det ekonomisk-politiska området (*Nationalräkenskapsdata för Åland 1993 – 1996*).

1.2 Bakgrund

Åland tillhör Finland men har sedan 1921 ett vidsträckt självstyre. Landskapet består av 6 500 öar och skär varav den största ön fasta Åland utgör ungefär 70 procent av den totala landarealen. Från norr till söder är fasta Åland 50 km och från öster till väster 45 km (*Åland kort och gott*, 2002). Den enda egentliga staden på Åland, Mariehamn, utgör landskapets administrativa och ekonomiska centrum. Det bor cirka 26 000 människor på Åland, varav ungefär 40 procent i Mariehamn. Högsta exekutiva instans på Åland är Landskapsstyrelsen (*Holmertz*, 1997).

Från att ha varit ett utpräglat jordbrukssamhälle har Ålands huvudnäringar övergått till att vara sjöfart samt sedan 1960-talet även turism. En viktig anledning till turismens ökade betydelse var färjetrafikens expansion. I och med snabba, billiga färjeturer från bland annat Stockholm och Helsingfors blev Åland ett populärt turistmål.

Dock finns det fortfarande problem med det isolerade läget. Jordbruk och industri lider konkurrensmässiga nackdelar på grund av höga transportkostnader för tunga och skrymmande varor. Jordbruket har därför tenderat att specialisera sig på vissa nischprodukter. Som exempel kan nämnas så kallade trädgårds- och grönsaksprodukter (lök, gurka, tomater, sallad, äpplen, dill mm) vilka tillsammans med sockerbetsproduktion år 2001 stod för 42 procent av det åländska jordbrukets bruttointäkter. De mer klassiska jordbruksprodukterna såsom råg, vete, korn, havre och oljeväxter stod samma år för endast fem procent av jordbrukets bruttointäkter (*Lindström m.fl.*, 2003). Industrin har till stor del specialiserat sig som underleverantörer till utomåländsk storindustri samt på produktion av livsmedel och plastprodukter.

Ålands självstyrelse gör att behovet av kvalificerade uppgifter om landskapets ekonomi är stort. Åland för därför egna nationalräkenskaper, och ÅSUB är officiell producent av dessa (*Nationalräkenskapsdata för Åland 1993 – 1996*).

Ett av de viktigaste måtten på en regions ekonomi är bruttonationalprodukten (BNP). Generellt finns det tre angreppssätt för att beräkna BNP: från inkomstsidan, från utgiftssidan samt enligt förädlingsvärdesprincipen. Ålands BNP beräknas enligt förädlingsvärdesprincipen. BNP enligt förädlingsvärdesprincipen kan delas upp i det privata näringslivet, offentlig verksamhet samt icke-vinstsyftande privat verksamhet. Förädlingsvärdet för ett företag i det privata näringslivet definieras som differensen mellan ett företags omsättning och värdet av dess insatsförbrukning (*Nationalräkenskapsdata för Åland 1993 – 1996*). Genom att summera förädlingsvärdet för samtliga företag fås det privata näringslivets tillskott till BNP.

1.3 Syfte och problemformulering

Förädlingsvärdet för företagen på Åland kan beräknas med hjälp av uppgifter om omsättning och insatsförbrukning från Skattestyrelsens (Skattemyndighetens) årliga bokslutsregister. Dock blir dessa uppgifter tillgängliga först med ett eller flera års fördröjning. För att kunna ta ekonomiska beslut är det viktigt att snabbt kunna uppdatera informationen om BNP. Idag använder man därför olika sekundära indikatorer för att få en uppskattning av BNP innan uppgifterna från bokslutsregistret finns tillgängliga. Hur detta bäst bör ske har emellertid inte varit föremål för någon systematisk studie.

Förutom det årliga bokslutsregistret publicerar Skattestyrelsen ett månatligt momsregister, som vanligen blir tillgängligt med två till tre månaders fördröjning. I momsregistret finns ej samma uppgifter som i bokslutsregistret men det innehåller ändå uppgifter som kan antas ha hög korrelation med förädlingsvärdet.

Syftet med detta projektarbete är att utveckla en metod för att beräkna det privata näringslivets tillskott till BNP med hjälp av de uppgifter som framkommer ur det månatliga momsregistret. Metoden ska användas för att prognostisera de årliga procentuella förändringarna av det privata näringslivets tillskott till BNP innan uppgifterna från bokslutsregistret är kända. Tack vare att momsregistret uppdateras så ofta skulle detta ge en mycket snabb indikator av det privata näringslivets tillskott till BNP.

1.4 Omfattning

Projektet har av uppdragsgivaren avgränsats till att omfatta det privata näringslivet. Beräkningarna har baserats på data för år 1996 – 2001.

1.5 Disposition

Efter detta inledande kapitel följer i kapitel två en beskrivning av datamaterialet som projektet baserats på samt preliminär bearbetning av detta. Kapitel tre leder i diskuterande ordalag fram till en teoretiskt lämplig modell som kan användas för prognostisering. I kapitel fyra anpassas denna modell till föreliggande data och modellens parametrar bestäms. I kapitel fem tillämpas modellen för att beräkna det privata näringslivets tillskott till BNP för år 2000 och 2001. Förfarandet kan liknas vid prognostisering eftersom de värden som beräknas ej fanns publicerade vid tidpunkten för projektets genomförande. Kapitel sex beskriver en alternativ metod som enbart baseras på de tio företag som har störst påverkan på det privata näringslivets summerade förädlingsvärde. Kapitel sju knyter på ett sammanfattande sätt ihop slutsatserna från metod och resultat. I detta kapitel redovisas även författarens egna slutsatser och rekommendationer.

2. Datamaterial och preliminär bearbetning

I detta kapitel beskrivs det tillgängliga datamaterialet samt resultaten av den preliminära bearbetningen av dessa data. Dessutom förs en diskussion kring felkällor i datamaterialet.

2.1 Konstruktion

Data har inhämtats från tre olika källor.

Statistikcentralens (Finlands statistiska centralbyrå) *företagsregister* användes för att länka samman övriga register. Detta var möjligt genom företagens unika FO-nummer (företagsorganisationsnummer). Registret innehöll data för år 1996 – 2001. Från registret hämtades även information om företagens branschtillhörighet, vilken kommun de var verksamma i samt vilka år de existerat.

Skattestyrelsens *bokslutsregister* innehåller de officiella uppgifterna om företagen. Vid tidpunkten för konstruktionen av datamaterialet fanns data för år 1996 – 1999. Från bokslutsregistret hämtades uppgifter om företagens förädlingsvärden, socialavgifter, insatsförbrukning, omsättning samt lönekostnader.

Skattestyrelsens månatliga *momsregister* innehåller de uppgifter som kan användas som förklarande variabler vid skattning av förädlingsvärdet och som bas för prognostiseringen. Data fanns tillgängliga för år 1996 – 2001. Från momsregistret hämtades uppgifter om företagens lönekostnader, omsättning, socialavgifter, handel med andra EU-länder samt momsuppgifter.

De variabler som var av intresse för projektet ställdes av uppdragsgivaren samman i ett gemensamt datamaterial. De månatliga uppgifterna från momsregistret summerades årsvis. Det sammanställda datamaterialet kom att bestå av 2021 företag under år 1996 – 2001. I *Bilaga I* finns en förteckning med beskrivningar av samtliga variabler.

2.2 Beskrivning

2.2.1 Datatyp

Data är av typen individdata, där individerna utgörs av företag. Generellt gäller att om man vid flera tidpunkter studerar en individ med avseende på en eller flera egenskaper erhåller man tidsseriedata. Om man vid en enda tidpunkt studerar en eller flera individer med avseende på en eller flera egenskaper erhåller man tvärsnittsdata. Om man kombinerar tidsseriedata och tvärsnittsdata, det vill säga att man vid flera tidpunkter studerar samma individer med avseende på en eller flera egenskaper, erhåller man paneldata. För detta projekts data har pa-nelen tre dimensioner: tid, bransch och företag.

Datamaterialet innehåller årliga uppgifter om 2 021 företag på Åland under tidsperioden 1996 – 2001. Viktigt att observera är att datamaterialet trots den intentionen ej kan göra anspråk på att innehålla alla företag på Åland under den aktuella tidsperioden. Detta berodde på flera faktorer. Vid sammanställning av datamaterialet använde ÅSUB tre olika register. Kopplingar mellan dessa register skedde med hjälp av företagens FO-nummer. Dock har det under den aktuella tidsperioden 1996 – 1999 skett en förändring i FO-numren. I början användes för enmansföretag näringsidkarens personnummer som FO-nummer. Senare övergick man till att alla företag oavsett storlek skulle ha ett speciellt åttasiffrigt FO-nummer. Vid sammanlänkning av registren fick därför ytterligare ett register med kopplingar mellan FO-nummer och personnummer användas. I detta register saknades vissa företag som därför kom att försvinna ur det slutliga datamaterialet. Många småföretag försvann till följd av att företag med omsättning mindre än 8 500 euro/år ej betalar månatlig moms och därmed inte finns med i det månatliga momsregistret. I vissa fall saknas observationer till följd av att företagets bokslutsperiod ej sammanfaller med kalenderåret.

Trots att det slutliga sammanställda datamaterialet innehöll 2 021 (oidentifierade) företag fanns det inte uppgifter för exakt 2 021 företag vid samtliga dessa tidpunkter. Detta beror främst på att vissa företag bildats under undersökningsperioden medan andra företag lagts ned. Av de totalt 12 156 observationerna saknades därför 1 314 värden. Olika former av imputation av saknade värden övervägdes, men förkastades eftersom bortfallet inte var slumpmässigt. Dessutom konstaterades att de företag för vilka data delvis saknades svarade för en mycket liten andel av det totala BNP-värdet.

Det erhållna datamaterialet kan beskrivas som ett obalanserat paneldatamaterial. Företagsdata har vidare den speciella egenskapen att företag naturligt grupperas i branscher. Företag inom samma bransch kan förväntas uppvisa mer likartade egenskaper än företag i övriga branscher och påverkas dessutom mer likformigt av yttre påverkan, exempelvis konjunktursvängningar.

2.2.2 Numerisk beskrivning

Datamaterialet innehöll tre olika bransch-koder med åtta, tio respektive 349 olika klasser. I samråd med uppdragsgivaren beslutades att genomgående utnyttja bransch-koden med tio olika klasser. I tabell 2.1 redovisas medelvärden och standardavvikelser för förädlingsvärdena branschvis under år 1996 – 1999. Inom parentes i respektive ruta anges även hur många aktiva företag som fanns i branscherna varje år. Varje bransch har studerats vid alla de fyra åren 1996 – 1999. Att det finns fyra tidsobservationer för varje bransch trots att inte varje företag funnits varje år följer av att branscherna är så stora att åtminstone ett företag existerat varje år i varje bransch.

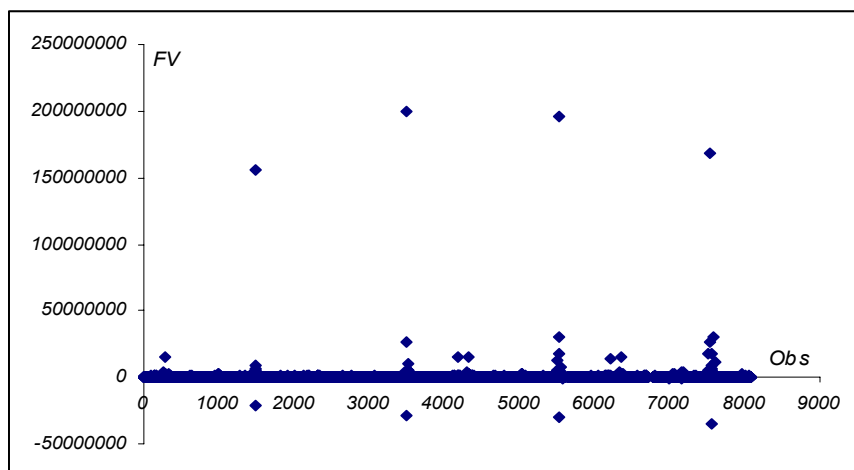
Tabell 2.1 Medelvärden och standardavvikelser för branschernas förädlingsvärden (tusental euro) samt antal företag i respektive bransch, fördelat per år

Bransch	1996	1997	1998	1999
Primärnäringsar	71; 163 (77)	65; 162 (97)	85; 251 (98)	96; 253 (94)
Livsmedelsindustri	297; 448 (22)	95; 251 (27)	711; 2772 (30)	689; 2450 (30)
Övrig industri	313; 1530 (121)	89; 295 (146)	278; 1354 (150)	288; 1425 (144)
Vatten- och elproducenter	450; 834 (11)	209; 526 (12)	511; 691 (13)	469; 666 (13)
Byggverksamhet	49; 134 (254)	35; 83 (283)	69; 168 (299)	79; 170 (290)
Handel	88; 263 (381)	53; 131 (428)	106; 299 (429)	119; 377 (410)
Hotell och restaurang	80; 188 (102)	31; 77 (133)	82; 187 (139)	96; 221 (139)
Transp. och kommunikation	1108; 12134 (170)	1302; 14831 (190)	1431; 14310 (199)	1365; 12737 (192)
Bank och försäkring	186; 420 (17)	46; 143 (21)	288; 760 (25)	1700; 6313 (26)
Övriga tjänster	50; 139 (318)	26; 62 (388)	54; 163 (420)	49; 148 (432)

De största branscherna sett till antalet företag är *Handel* och *Övriga tjänster*. Det högsta förädlingsvärdet finns dock inom branschen *Transport och kommunikation*. Samtliga branscher utom *Transport och kommunikation* hade en kraftig svacka år 1997. Svackan kan ej förklaras med att företag lagts ned, då antalet företag i samtliga branscher ökade 1997. Svackan verkar heller inte ha fått företag att gå i konkurs, då antalet företag i varje bransch ökade till år 1998. Enligt uppdragsgivaren är orsaken till den kraftiga svackan att många företag på grund av fel vid konstruktion av registren försvunnit år 1997. Samtliga branscher utom *Övrig industri* och *Övriga tjänster* har ökat sitt förädlingsvärde under perioden 1996 – 1999. Branschen *Bank och försäkring* står för den procentuellt sett största ökningen.

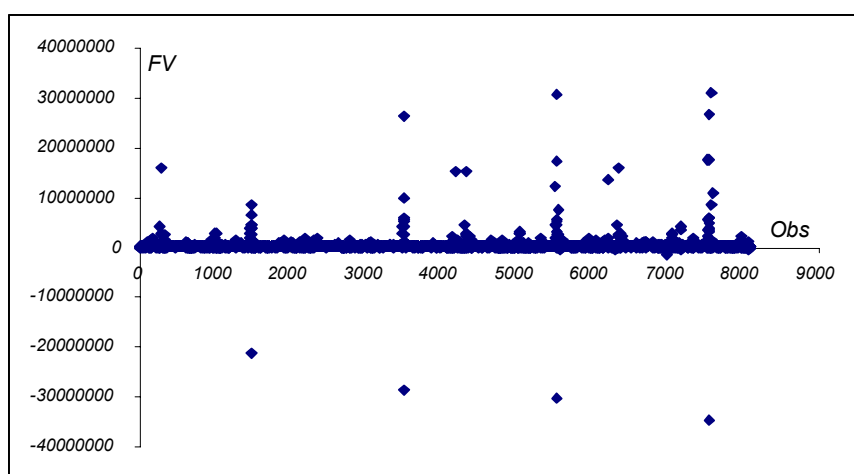
Figur 2.1 åskådliggör förädlingsvärdena för samtliga observerade företag. Företagen har sorterats efter år och sedan efter bransch. Observera att varje företag därmed har plottats minst en gång och högst fyra gånger, beroende på hur många av de observerade åren som företaget existerat.

Fig 2.1 Förädlingsvärde för varje företag, observationerna ordnade årsvis (euro)



Som framgår av figur 2.1 finns fyra extremt avvikande observationer. Dessa tillhör samma företag, ett företag i branschen *Transport och kommunikation*. I syfte att förtydliga variationen i förädlingsvärde för övriga företag skapades ett filter som sorterade bort detta företags observationer.

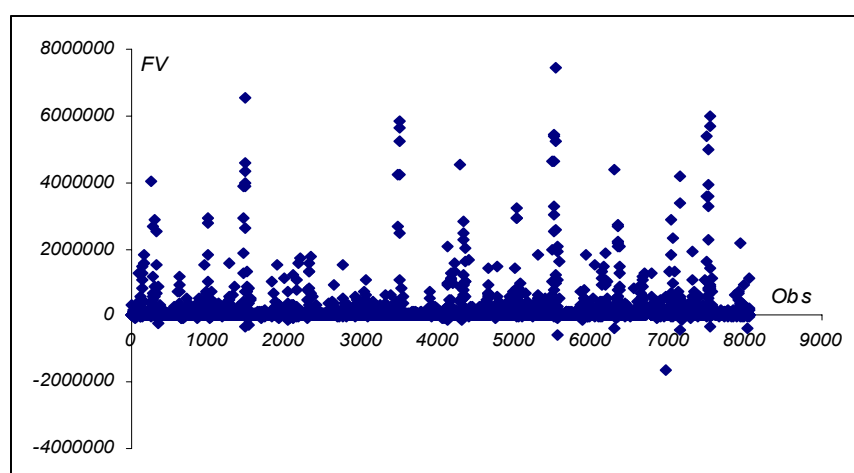
Fig 2.2 Förädlingsvärde för varje företag, observationerna ordnade årsvis, filtrerat (euro)



Därefter framträder ytterligare ett antal observationer med kraftigt avvikande värden. Särskilt utmärker sig fyra observationer med extremt negativa förädlingsvärden. Observationerna kommer från samma företag, och även detta företag tillhör branschen *Transport och*

kommunikation. Den generella förklaringen till negativa förädlingsvärden är att företaget ingår i en koncern. Inköp av insatsvaror och omsättning bokförs inom olika delar av koncernen, och därmed får den del av koncernen som enbart hanterar inköp negativa förädlingsvärden. Negativa förädlingsvärden kan även bero på att företaget byggt upp ett lager. Ett nytt filter skapades för att sortera bort dessa negativa observationer samt alla observationer med förädlingsvärden större än 8 000 000 euro. Observera att gränsen 8 000 000 euro valdes subjektivt, men därför att det vid denna nivå framträdde en naturlig gräns.

Fig 2.3 Förädlingsvärde för varje företag, observationerna ordnade årsvis, filtrerat två gånger (euro)



I figur 2.3 har sammanlagt 25 observationer tillhörande tio olika företag filtrerats bort. Bland kvarvarande observationer finns förvisso sådana som avviker men ingen som avviker så kraftigt som de som filtrerats bort. Slutsatsen blir att det åländska näringslivet består av många jämnstora företag och ett tiotal mycket stora företag.

2.2.3 Fördelning

Många statistiska metoder baseras på att modellens förklarande variabler är normalfördelade. Två kontroller för normalfördelning gjordes därför. Dels skapades normalfördelningsplottar, och dessutom utfördes Kolmogorov-Smirnoffs test. Detta test mäter den största absoluta skillnaden mellan observerade och förväntade värden under nollhypotesen att data uppvisar normalfördelning. Normalfördelningsplottarna var för samtliga variabler skeva åt höger. För samtliga variabler förkastades även nollhypotesen om normalfördelning vid Kolmogorov-Smirnoffs test. Bägge testen indikerar alltså att data ej uppvisar normalfördelning. Två orsaker till detta kunde härledas. Den första och viktigaste orsaken följer ur företagsstrukturen på Åland där ett litet antal mycket stora företag ger kraftigt avvikande värden. Den andra orsaken var att många variabler innehöll en hög frekvens nollor. Detta följer ur att många företag haft nollvärden för dessa variabler, men i vissa fall verkade nollorna även bero på felaktiga uppgifter i registren. En vanlig åtgärd när observationer ej är normalfördelade är att

transformera data. Gemensamt för majoriteten transformationsmetoder är att de går ut på att pressa ned stora och avvikande observationer och på så sätt få data som bättre följer normalfördelningskurvan. I det aktuella projektet är detta ingen lämplig metod. Eftersom de mycket stora företagen har stor inverkan på summan av de åländska företagens förädlingsvärde bör de också ha stor ef-fekt på parametrarna i en prognosmodell.

Avvikelserna från normalfördelningen kräver speciell omsorg vid valet av skattningsmetod och försvarar beräkningar av osäkerhet.

2.3 Bearbetning

På uppdragsgivarens inrådan inkluderades bransch nummer fyra, *Vatten- och elproducenter*, i bransch nummer tre, *Övrig industri*. Anledningen var dels att bransch nummer fyra innehöll så få företag att det ansågs onödigt att bilda en egen grupp för denna och dessutom att den passade bra in i *Övrig industri*.

De variabler som hämtats från bokslutsregistret var i valutan finska mark medan variabler som hämtats från momsregistret var i valutan euro. Eftersom euro sedan första januari 1999 är officiell valuta i Finland räknades bokslutsregistrets uppgifter om till euro. I samband med Finlands övergång från mark till euro fixerades växelkursen 5,94573 mark/euro. Denna växelkurs användes som omräkningsfaktor.

2.4 Felkällor i datamaterialet

Momsregistret utgör en kvalitetsmässigt ojämn källa av data för vissa branscher. Branschen *Primärnärningar* innehåller många företag som av diverse anledningar betalar moms årsvis, och dessa företag finns därmed inte med i det månatliga momsregistret. I vissa branscher är antalet observationer för få för att de branschvisa resultaten ska vara helt tillförlitliga. Trots dessa problem har branschvisa resultat redovisats då de är betydelsefulla för förståelsen av uppbyggnaden av det privata näringslivets tillskott till BNP och prognosmetoden.

Under arbetets gång upptäcktes vissa fel i datamaterialet. Bokslutsregistret och momsregistret innehöll båda variabler med löner respektive omsättning. Efter omräkning till gemensam valuta borde lönevariablerna respektive omsättningsvariablerna från vardera registret sammanfalla, fränsett mindre avvikelser till följd av avrundningsfel. Dock rädde det för vissa företag mycket stora differenser mellan de två registren. För en del företag upptäcktes dessutom fel såtillvida att vissa uppgifter om företagen saknades samt att nollor ibland felaktigt satts som variabelvärden. Anledningen till felen lokaliserades till bristande kontrollrutiner vid Skattestyrelsens framställning av registren. På uppdragsgivarens inrådan prövades att skapa ett filter som sorterade bort observationer vars uppgifter i bokslutsregistret

skilde sig mer än en viss procent från uppgifterna i momsregistret. Olika procentsatser prövades men redan vid gränsen ± 40 procent försvann mer än 3 000 observationer ur registret. Bortfallet ansågs för stort för att filtret skulle kunna användas. Detta innebär att de felaktiga inmatningarna kan utgöra en all-varlig felkälla. I samråd med uppdragsgivaren beslutades ändå att acceptera den felkälla som de felaktiga registerinmatningarna kan medföra. Data från år 1999 och framåt har underkastats noggrannare kontrollrutiner vilket borgar för bättre kvalitet på data i framtiden.

Att observationer saknas i materialet utgör givetvis en allvarlig felkälla. I kapitel 2.2.1 framgick tre viktiga orsaker till att observationer försvunnit: ett nytt system för FO-nummer som gjorde att observationer försvann vid sammanlänkning av registren, småföretag som ej betalar månatlig moms och därmed inte finns med i momsregistret samt företag vars bokslutsår inte sammanfaller med kalenderåret. Särskilt många observationer har försvunnit ur materialet år 1997. Detta kommer givetvis att påverka prognosernas kvalitet. I syfte att undersöka hur stor inverkan saknade observationer kan tänkas ha på parametrarna i en prognostiseringsmodell beräknades det genomsnittliga förädlingsvärdet, dels för de företag som inte existerat varje år, dels för de företag vars observationer på grund av fel något eller några år försvunnit ur analysen.

Tabell 2.2 Genomsnittligt förädlingsvärde för olika kategorier av företag (euro)

Företag	Ej existerat alla år	Saknat ett eller flera år	Alla företag
FV	188 348,1	152 294,5	235 689,4

Som framgår ur tabell 2.2 är de genomsnittliga förädlingsvärdena betydligt lägre för de grupper av företag som inte existerat alla år eller vars observationer försvunnit ett eller flera år. Detta innebär att de saknade företagen utgör en begränsad felkälla. I samråd med uppdragsgivaren beslutades att acceptera denna felkälla, då det är den teoretiska framställningen av metoden som för uppdragsgivaren är av primärt intresse.

3. Metod

Detta kapitel leder fram till en modell som är teoretiskt lämplig för den typ av data som framkommer ur moms- och bokslutsregistret. Dessutom beskrivs hur modellen kan skattas samt hur den kan användas för prognostisering.

3.1 Modell

Projektet syftar till att bestämma en metod som beskriver förädlingsvärdet för det privata näringslivet på Åland som en funktion av relevanta uppgifter från det månatliga momsregistret. Metoden ska användas för att prognostisera de årliga procentuella förändringarna av det privata näringslivets tillskott till BNP.

Detta är möjligt genom att till data anpassa en regressionsfunktion med förädlingsvärde som beroende variabel. För år 1996 – 1999 finns data från både bokslutsregistret och momsregistret. Dessa data kan användas för att bestämma modellens parametrar. Momsregistret innehåller dessutom data för år 2000 och 2001. Genom att sätta in dessa data i regressionsfunktionen fås prognoser av förädlingsvärdet för år 2000 och 2001. På detta sätt kan nya prognoser av förädlingsvärdet tas fram varje gång momsregistret uppdateras. När bokslutsregistret uppdateras är det lämpligt att skatta om modellens parametrar. Modellen kommer därigenom successivt att förbättras. Detta ställer valet av modell i en ganska speciell situation: av primärt intresse är snarare att bestämma en modell som passar till den typ av data som framkommer ur moms- och bokslutsregistren, än att bestämma en modell som passar hundraprocentigt till just det datamaterial som fanns tillgängligt vid tidpunkten för projektets genomförande.

Ett obalanserat paneldatamaterial bestående av företagsdata med naturliga grupperingar kräver speciell omsorg vid valet av modell. Paneldata kräver att modellen kan hantera datamaterialets olika dimensioner, i det här fallet tid, bransch och företag. I ett paneldatamaterial över företagsdata finns två källor till korrelation som modellen måste kunna hantera. De naturliga grupperingarna innebär risk för korrelation mellan företag i samma bransch. Dessutom finns en tidsdimension i datamaterialet, eftersom varje företag har observerats vid flera tidpunkter. Detta innebär risk för korrelation mellan varje observation av respektive företag (*Baltagi, 1999*). Att panelen är obalanserad ställer krav på valet av skattningsmetod.

Inom paneldatamodellering finns det två modellkategorier att utgå ifrån (*Gujarati, 1995*). Dessa kategorier betecknas *fixed effect models* (FEM) samt *random effect models* (REM). I en modell av typen FEM hanteras paneldatastrukturen genom indikatorvariabler. Indikatorvariablerna (vanligtvis dikotoma) används för att beskriva observationernas tillhörighet i tid och rum. Detta skulle för de för projektet föreliggande data kräva 2020

indikatorvariabler för att hantera företagsdimensionen samt ytterligare åtta indikatorvariabler för att hantera branschdimensionen i data. I en modell av typen REM hanteras paneldatastrukturen istället genom att varje observations tillhörighet i tid och rum inkluderas som olika komponenter i den totala slumpvariationen. Detta skulle i det aktuella fallet innebära att det krävs en komponent för branscherna, en komponent för företagen, samt en komponent för den återstående slumpstörningen. Det finns inga exakta regler för att avgöra när FEM respektive REM bör användas. Vanligen används dock FEM när responsvariabeln påverkas av ett måttligt antal faktorer och de olika faktorernas nivåer är givna, medan REM ursprungligen utvecklades för att hantera slumpmässigt bestämda faktornivåer. I praktiken har REM också kommit att användas när paneldatastrukturen har många dimensioner eller om det finns många källor till korrelation i data. Det går naturligtvis även att tänka sig kombinationer av FEM och REM.

Inom statistiken finns en allmän strävan efter att beskriva data med så få parametrar som möjligt. En modell med fixa effekter kräver i detta sammanhang 2032 parametrar medan en modell med slump effekter endast kräver sju parametrar. Detta talar för att använda en modell av typen REM. Generellt gäller även att REM passar bättre om panelen är obalanserad (*Baltagi m.fl.*, 2001). Speciellt brukar man under dessa omständigheter använda mixade linjära modeller (*Milliken*, 2002), som är generaliseringar av vanliga linjära modeller och speciellt utvecklade för att ge rimliga statistiska beskrivningar av slumpkomponenter som uppvisar beroende sinsemellan. En mixad linjär modell är uppbyggd av parametrar för fixa effekter samt av slump effekter. De fixa effekterna är de kända förklarande variabelernas påverkan på den beroende variabeln. Parametrarna för de fixa effekterna avser genomsnittsvärden för hela populationen. Slump effekterna är de individspecifika avvikelserna från det för alla individer gemensamma interceptet. Statistiskt sett är de slumpvariabler med väntevärde noll. Variansen för slump effekterna och samvariationen mellan olika slumpkomponenter brukar beskrivas med hjälp av kovariansparametrar.

Figur 3.1 visar att det krävs en $m \times m$ -matris med elementen $g_{ij} = \text{cov}(\gamma_i, \gamma_j)$ för att beskriva en kovariansstruktur för m slump effekter.

Fig 3.1 Generell kovariansstruktur

$$\text{Kovariansstruktur} = \begin{pmatrix} g_{11} & g_{12} & g_{13} & \dots & g_{1m} \\ g_{21} & g_{22} & g_{23} & \dots & g_{2m} \\ g_{31} & g_{32} & g_{33} & \dots & g_{3m} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ g_{m1} & g_{m2} & g_{m3} & \dots & g_{mm} \end{pmatrix}$$

Genom att identifiera de tänkbara källorna till korrelation i datamaterialet kan kovarianserna ofta beskrivas med ett mindre antal modellparametrar. Det finns många exempel på sådana förenklade kovariansstrukturer. Bland de vanligaste brukar nämnas autoregressiv struktur, ostrukturerat samt varianskomponentstruktur. I detta projekt kommer en varianskomponentstruktur att tillämpas.

I *Baltagi* (1999) beskrivs en mixad linjär modell som är speciellt utvecklad för obalanserade paneldata med naturliga grupperingar. Modellen, som betecknas *unbalanced nested error components regression model*, tillhör kategorin *random effect models* (REM). Modellen klarar av att hantera olika antal företag inom varje bransch samt olika antal observationer för varje företag. Låt μ_1, \dots, μ_n beteckna slumpeffekterna för branscherna $1, \dots, n$ och v_{11}, \dots, v_{nm} slumpeffekterna för företagen $1, \dots, m$ i branscherna $1, \dots, n$. Tack vare dessa slumpeffekter hanterar modellen förutom korrelationer även dimensionerna i data. Låt vidare p beteckna antalet fixa parametrar och q antalet slumpeffekter. Förutsatt att alla företag existerar vid alla tidpunkter finns det $t(m_1 + m_2 + \dots + m_n)$ observationer där t betecknar antalet observerade tidpunkter och m_1, \dots, m_n betecknar antalet företag i branscherna $1, \dots, n$. I praktiken har dock av orsaker som diskuterades i kapitel 2.2.1 inte alla företag studerats vid alla tidpunkter. Detta innebär att det verkliga antalet observationer kommer att vara något lägre, men betecknings sättet har av praktiska skäl ändå tillämpats. På matrisform kan modellen då tecknas enligt

$$\mathbf{y} = \mathbf{X}\boldsymbol{\alpha} + \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}$$

- y** Kolumnvektor av storlek $t(m_1 + m_2 + \dots + m_n) \times 1$ som innehåller de observerade förädlingsvärdena för alla företag vid alla tidpunkter
- X** Designmatris av storlek $t(m_1 + m_2 + \dots + m_n) \times (p + 1)$ som innehåller en kolumn ettor för interceptet och en kolumn vardera för de förklarande variablerna
- $\boldsymbol{\alpha}$** Vektor av storlek $(p + 1) \times 1$ med de okända fixa effekterna som modellen syftar till att skatta
- Z** Designmatris av storlek $t(m_1 + m_2 + \dots + m_n) \times q$ som innehåller indikatorvariabler för branscherna samt indikatorvariabler för företagen inom respektive bransch
- $\boldsymbol{\gamma}$** Vektor av storlek $q \times 1$ med de okända slumpeffekterna μ_1, \dots, μ_n samt v_{11}, \dots, v_{nm} som modellen syftar till att skatta
- $\boldsymbol{\varepsilon}$** Kolumnvektor av storlek $t(m_1 + m_2 + \dots + m_n) \times 1$ med de företagsspecifika återstående icke observerbara slumpstörningarna

På komponentform kan modellens tecknas

$$Y_{ijt} = \alpha_0 + \alpha_1 X_{1ijt} + \dots + \alpha_p X_{pijt} + \mu_i + v_{ij} + \varepsilon_{ijt}$$

där Y_{ijt} betecknar förädlingsvärdet för det j :te företaget i den i :te branschen vid tidpunkt t . $\alpha_0, \alpha_1, \dots, \alpha_p$ betecknar parametrarna för de fixa effekterna. μ_i är slumpeffekt för den i :te branschen och hanterar korrelation mellan företag i den i :te branschen. v_{ij} är slumpeffekt för det j :te företaget i den i :te branschen och hanterar korrelation till följd av att det j :te företaget observerats vid flera tidpunkter. ε_{ijt} betecknar den företagsspecifika återstående icke observerbara slumpstörningen för det j :te företaget i den i :te branschen vid tidpunkt t .

γ och ε antas vara oberoende och komponenterna i dess vektorer antas vara normalfördelade slumpvariabler med väntevärde noll och kovariansmatris \mathbf{G} respektive \mathbf{R} .

Om oberoendeantagandena förs ned på företagsnivå innebär de att slumpeffektkomponenterna μ_i , v_{ij} och ε_{ijt} antas vara oberoende. Vidare får slumpeffekterna ej uppvisa någon autokorrelation.

The unbalanced nested error components regression model har en kovariansstruktur som kan hantera korrelation mellan observationer av företag inom samma bransch samt korrelation till följd av att flera observationer gjorts av samma företag vid olika tidpunkter. Den första typen av korrelation modelleras inom \mathbf{G} eftersom de avser slumpeffekterna μ_i och v_{ij} (*SAS Users Manual*, kapitel 18). Enligt antagandena som ställdes upp ovan finns ingen korrelation mellan de två slumpeffekterna. Detta innebär att \mathbf{G} får en diagonal kovariansstruktur som brukar betecknas varianskomponentstruktur. Med beteckningarna i figur 3.1 innebär detta att alla kovariansparametrar utom $g_{11}, g_{22}, \dots, g_{mm}$ är noll. Kovariansparametrarna i huvuddiagonalen kommer däremot att ge en variansskattning vardera för respektive slumpeffektkomponent. σ_μ^2 blir variansen för den branschspecifika slumpeffekten och σ_v^2 blir variansen för den företagsspecifika slumpeffekten. \mathbf{R} hanterar korrelation inom den återstående slumpstörningen ε_{ijt} . Eftersom dessa antas vara oberoende får de utseendet $\mathbf{R} = \sigma_\varepsilon^2 \mathbf{I}$ där \mathbf{I} är enhetsmatrisen (*Baltagi m.fl.*, 2001). Variansen för den återstående slumpstörningen betecknas σ_ε^2 .

Modellexekvering resulterar i skattningar av parametrarna för de fixa effekterna, av parametrarna för slumpeffekterna samt av variansskattningar för respektive slumpeffekt.

3.2 Skattning

Skattning av en mixad linjär modell är mer komplicerat än skattning av en vanlig linjär modell. Förutom α ska även γ , \mathbf{G} samt \mathbf{R} skattas. Såvida σ^2_μ och σ^2_ν inte båda kan antas vara noll ger minsta kvadratskattning (OLS) ej väntevärdesriktiga skattningar. Istället krävs *generaliserad* minsta kvadratskattning (GLS) (*SAS Users Manual*, kapitel 18). Skillnaden är att GLS kan ta hänsyn till ytterligare information, exempelvis att det föreligger olika varians för olika grupper i data, och vikta observationernas betydelse för parameterskattningarna därefter. För en mixad linjär modell går GLS-metoden ut på att minimera

$$(\mathbf{y} - \mathbf{X}\alpha)' \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\alpha)$$

där $\mathbf{V} = \mathbf{ZGZ}' + \mathbf{R}$ och utgör variansen för \mathbf{y} . Problemet är att denna formel kräver information om \mathbf{V} , och därmed om kovariansparametrarna i \mathbf{G} och \mathbf{R} . Därför krävs först skattningar av \mathbf{G} och \mathbf{R} (*SAS Users Manual*, kapitel 18). I litteraturen (exempelvis *Wu m.fl.* 2001) framhålls tre olika skattningsmetoder som lämpar sig för skattning av kovariansparametrarna i mixade linjära modeller. Dessa är maximum likelihood (ML), restricted maximum likelihood (REML) samt minimum norm quadratic unbiased estimator (MINQUE). ML är en iterativ metod som i detta sammanhang baseras på att data är normalfördelade. Metoden går ut på att en likelihood-funktion maximeras med avseende på parametrarna av intresse. REML baseras på ML men delar upp likelihood-funktionen i två delar, varav den ena delen innehåller parametrar för fixa effekter och den andra delen parametrar för slump effekter. Även REML baseras på normal-fördelningsantagande. MINQUE är en ickeiterativ metod som baseras på momentskattning och som utvecklats speciellt för skattning av kovariansparametrarna i mixade linjära modeller. Metoden kräver inget fördelningsantagande, och det enda krav som ställs på data är att de första fyra momenten existerar. Metoderna är, sett till skattningarnas effektivitet och väntevärdesriktighet, tämligen likvärdiga (*Wu m.fl.*, 2001).

Statistikprogrammet SAS innehåller en procedur, Proc Mixed, som hanterar mixade linjära modeller. Proceduren klarar av samtliga ovan diskuterade skattningsmetoder.

3.3 Prognostisering

Modellvalsproceduren resulterar i en modell som beskriver förädlingsvärdet för varje företag som en funktion av de uppgifter som framkommer ur momsregistret. Bokslutsregistret uppdateras med ett par års fördröjning, medan momsregistret uppdateras med bara ett par månaders fördröjning. Genom att i modellen sätta in kända uppgifter från momsregistret av senare datum än uppgifterna från bokslutsregistret fås prognoser av varje företags förädlingsvärde fram till den sista observationen från momsregistret. Genom att summera de prognostiserade förädlingsvärdena för alla företag fås det privata näringslivets tillskott till

BNP. För detta projekts data leder dock raka summeringar till systematiska underskattningar av det privata näringslivets tillskott till BNP, på grund av att inte alla företag i det privata näringslivet ingår i analysen. Det privata näringslivets tillskott till BNP i absoluta tal kommer därför systematiskt att underskattas. Emellertid var projektets primära syfte att bestämma en metod för att prognostisera de *årliga procentuella förändringarna* av det privata näringslivets tillskott till BNP. Procentuella förändringar av det privata näringslivets tillskott till BNP beräknas genom att det summerade förädlingsvärdet för ett visst år divideras med det summerade förädlingsvärdet för föregående år. På så sätt reduceras den systematiska underskattningen, men givetvis saknas den påverkan som de företag som försvunnit ur analysen hade tillfört prognosen. Vid beräkning av de årliga procentuella förändringarna har, på uppdragsgivarens inrådan, endast de företag som existerar två på varandra följande år inkluderats. Logiken bakom detta är att uppgifter om många företag saknas i registret, och genom att endast ta med de företag som existerat två på varandra följande år minskar risken för att ett företag ska ha kraftig påverkan på det summerade förädlingsvärdet det ena året men saknas det därpå följande. Givetvis medför detta att observationer för företag som startat eller lagts ned under perioden kommer att uteslutas. Emellertid är det ett rimligt antagande att sådana företag inte är så stora att de har signifikant påverkan på resultaten.

Modellen baseras på ett datamaterial där de månatliga momsregisteruppgifterna har summerats årsvis. Trots detta krävs det inte ett helt års momsregisterdata för att en ny prognos ska kunna göras. Genom att summera momsregisteruppgifterna till och med det sista kända datumet kan man ändå använda prognostiseringsformeln, förutsatt att man håller reda på vilka månader prognosen gäller. Så snart momsregistret uppdaterats kan en ny prognos göras.

Prognoserna bygger på antagandet att det förhållande mellan förädlingsvärde och modellens förklarande variabler som gällt under perioden som parameterskattningarna baserats på kommer att fortsätta gälla under perioden som prognostiseras. När bokslutsregistret uppdateras framkommer ny information om förhållandet mellan modellens parametrar, och därmed är det lämpligt att skatta om modellens parametrar. På detta sätt kommer modellen med tiden att förbättras. Därigenom kommer den också att ge bättre och bättre prognoser.

4. Utförande

I föregående kapitel presenterades en teoretiskt lämplig modell. Nästa steg blir att anpassa och utveckla denna modell så att den optimeras till för projektet föreliggande data.

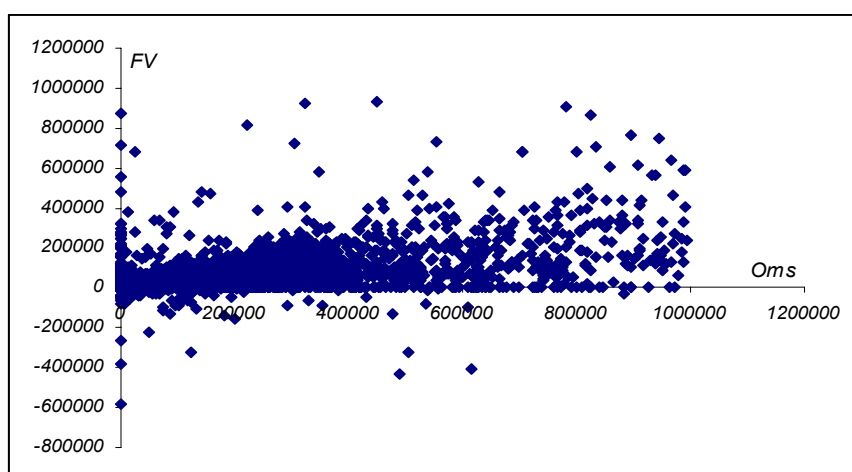
4.1 Förklarande variabler

Modellens förklarande variabler ska hämtas från det (i teorin) månatliga momsregistret. Detta register innehåller uppgifter om företagens lönekostnader, omsättning, socialavgifter, handel med andra EU-länder samt momsuppgifter. Vid konstruktion av en modell är det viktigt att de förklarande variablerna förutom att uppvisa ett faktiskt samband med den beroende variabeln även har teoretisk förankring till denna. Nedan följer en diskussion kring vilka av momsregistrets variabler som bör ingå i modellen.

Genom att plotta den presumtiva förklarande variabeln mot den beroende variabel framgår om det föreligger något samband mellan variablerna. För att öka tydligheten i figurerna skapades filter som sorterade bort observationer vars förädlingsvärde var större än $\pm 1\,000\,000$ euro, omsättning större än $1\,000\,000$ euro, moms till godo från alla inköp större än $400\,000$ euro eller socialavgiftsinbetalningar större än $400\,000$ euro. Observera att gränserna valdes subjektivt men på ett sådant sätt att relevanta samband framkom. För varje figur framgår hur många observationer som filterats bort.

Förädlingsvärde definieras i de åländska nationalräkenskaperna som differensen mellan ett företags omsättning och värdet av dess insatsförbrukning. Sett till förädlingsvärdets definition är det därför ett rimligt antagande att företagets förädlingsvärde är korrelerat med dess omsättning.

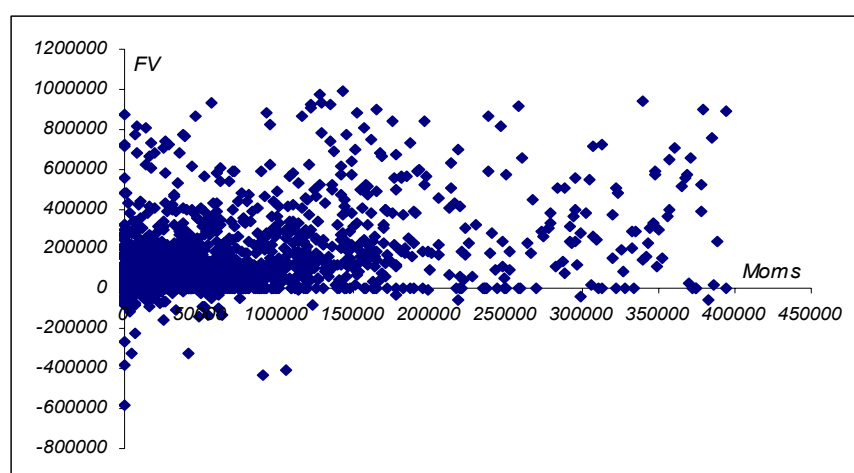
Fig 4.1 Sambandet mellan förädlingsvärde och Momsoms



Figur 4.1 åskådliggör sambandet mellan förädlingsvärde och omsättning (*Momsoms* enligt terminologin i *Bilaga I*). 647 observationer har filterats bort. Diagrammet bekräftar antagandet om att förädlingsvärde är korrelerat med omsättning. Såväl på teoretisk som praktisk basis bör *Momsoms* ingå som förklarande variabel i modellen.

I momsregistret framkommer inga direkta uppgifter om företagens insatsförbrukning. Men eftersom företagen betalar moms på alla inköp borde företagets momsuppgifter vara goda indikatorer på insatsförbrukningen. I momsregistret finns tre momsvariabler, moms till godo från alla inköp, moms som ska betalas in för företagets inhemska försäljning samt moms som faktiskt betalats in. Variabeln med moms som företagen faktiskt betalat in utgår från företagets försäljning och har sedan justerats för företagets inköp. Detta gör variabeln till en osäker indikator av företagets inköp. Variabeln med moms som ska betalas in för företagets inhemska försäljning ger inga direkta uppgifter om företagets inköp. Av dessa anledningar beslutades att utgå från variabeln moms till godo från alla inköp (*Momstg*), eftersom denna innehåller uppgifter om alla inköp, såväl inom som utom landet.

Fig 4.2 Sambandet mellan förädlingsvärde och Momstg

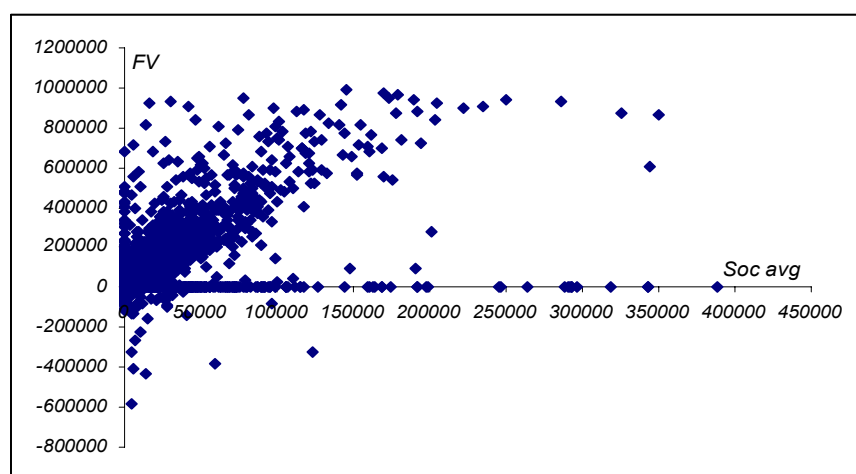


Figur 4.2 åskådliggör sambandet mellan förädlingsvärde och *Momstg*. 345 observationer har filterats bort för att förtydliga eventuella mönster i det återstående datamaterialet. Observationerna faller emellertid i hela kvadranten och inget entydigt mönster framgår. Variabeln har följaktligen god teoretisk men något osäker praktisk koppling till den beroende variabeln. Den teoretiska förankringen är dock så stark att variabeln bör prövas i modellen, för att se om den tillför modellen ökad förklaringsgrad. Sett till förädlingsvärdets definition kan parameterskattningen förväntas vara negativ.

Övriga uppgifter i momsregistret med teoretisk koppling till förädlingsvärdet avser företagets lönekostnader. Även löneuppgifter kan förväntas tillföra modellen ökad förklaringsgrad. Detta följer återigen ur förädlingsvärdets definition. Genom att utgå från företagets omsättning och subtrahera dess insatsförbrukning kvarstår förädlingsvärdet, och förädlingsvärdet består då grovt räknat av företagets lönekostnader samt dess vinst. Härur

följer att det borde finnas ett samband mellan förädlingsvärdets storlek och företagets lönekostnader. Momsregistret innehåller uppgifter om företagens faktiska lönekostnader samt uppgifter om företagens socialavgifter. I ett första steg bör givetvis lönevariabeln prövas som förklarande variabel. Men eftersom företagen betalar socialavgift på löner men ej på någonting annat borde uppgifter om företagets socialavgifter innehålla ungefär samma information. Prövning visade på bättre samband mellan förädlingsvärde och företagets faktiskt inbetalda socialavgifter (*Momssocbet*) än mellan förädlingsvärde och löner.

Fig 4.3 Sambandet mellan förädlingsvärde och Momssocbet



Figur 4.3 åskådliggör sambandet mellan förädlingsvärde och *Momssocbet*. 249 observationer har filtrerats bort. I diagrammet framgår att ett samband föreligger mellan variablerna. *Momssocbet* bör inkluderas i modellen.

En korrelationsmatris skapades i syfte att undersöka risken för multikollinearitet i modellen. I korrelationsmatrisen har dessutom den beroende variabeln förädlingsvärde (*SkattFV*) inkluderats. Matrisen har baserats på data för år 1996 – 1999.

Tabell 4.1 Korrelationsmatris för modellens presumtiva förklarande variabler samt den beroende variabeln

	<i>SkattFV</i>	<i>Momsoms</i>	<i>Momstg</i>	<i>Momssocbet</i>
<i>SkattFV</i>	1	0,954	0,091	0,705
<i>Momsoms</i>	0,954	1	0,262	0,790
<i>Momstg</i>	0,091	0,262	1	0,344
<i>Momssocbet</i>	0,705	0,790	0,344	1

Korrelationerna mellan *Momsoms* och *Momstg* samt mellan *Momstg* och *Momssocbet* är försumbara. Däremot är korrelationen mellan *Momsoms* och *Momssocbet* något högre. Detta är förvisso inte särskilt förvånande: företag med hög omsättning har oftast många anställda vilket leder till höga lönekostnader och därmed höga socialavgifter. Åtgärder mot eventuell multikollinearitet i modellen diskuteras i kapitel 4.5. I matrisen framgår som väntat mycket

god korrelation mellan *SkattFV* och *Momsoms*. Även korrelationen mellan *SkattFV* och *Momssocbet* är god. Däremot är korrelationen mellan *SkattFV* och *Momstg* anmärkningsvärt låg. Dock är den teoretiska kopplingen mellan förädlingsvärde och *Momstg* så pass starkt att det ändå kan anses motiverat att ta med variabeln i modellen.

Slutsatsen blir således att variablerna *Momsoms*, *Momstg* samt *Momssocbet* bör prövas som förklarande variabler.

Hittills har det förutsatts att alla företag kan beskrivas med en gemensam lutningsparameter. En naturlig utveckling av modellen blir att ifrågasätta detta antagande. Genom att skapa indikatorvariabler för varje företag och multiplicera dessa med modellens förklarande variabler tillåts företags specifika avvikelser från den gemensamma lutningsparameter.

4.2 Modellval

I tabell 4.2 redovisas vilka AIC-värden (Akaikes Information Criterion), som kan uppnås med olika huvudtyper av modeller. I princip anses en modell förklaringsvärde vara bättre ju lägre AIC-värdet är (*Makridakis m.fl.*, 1998). Till skillnad från MSE (summan av residualkvadratfele) som alltid sjunker när fler förklarande variabler införs i modellen, innehåller AIC-värdet en bestraffning av modeller med många parametrar. I denna studie användes AIC-värdet för att grovsålla bland modellerna. Sedan har modellerna med lägst AIC-värden granskats närmare.

Samtliga modellexekveringar har gjorts i SAS-proceduren Proc Mixed. Exempel på programkod återfinns i Bilaga II. Som skattningsmetod har MINQUE samt GLS använts.

Tabell 4.2 Några intressanta huvudtyper av modeller för beskrivning av modellvals-förfarandet

Nr	Modell	AIC
1	$Y_{ijt} = \alpha_0 + \alpha_1 X_{1ijt} + \varepsilon_{ijt}$	210604,7
2	$Y_{ijt} = \alpha_0 + \alpha_1 X_{1ijt} + \mu_i + v_{ij} + \varepsilon_{ijt}$	207857,3
3	$Y_{ijt} = \alpha_0 + \alpha_1 X_{1ijt} + \alpha_2 D_1 X_{1ijt} + \dots + \alpha_{2020} D_{2020} X_{1ijt} + \mu_i + v_{ij} + \varepsilon_{ijt}$	191260,2
4	$Y_{ijt} = \alpha_0 + \alpha_1 X_{1ijt} + \alpha_2 X_{2ijt} + \mu_i + v_{ij} + \varepsilon_{ijt}$	207133,4
5	$Y_{ijt} = \alpha_0 + \alpha_1 D_1 + \dots + \alpha_8 D_8 + \alpha_9 X_{1ijt} + \alpha_{10} X_{2ijt} + \alpha_{11} X_{3ijt} + v_j + \varepsilon_{jt}$	206554,7
6	$Y_{ijt} = \alpha_0 + \alpha_1 X_{1ijt} + \alpha_2 X_{2ijt} + \alpha_3 X_{3ijt} + \mu_i + v_{ij} + \varepsilon_{ijt}$	206757,8

X_{1ijt} omsättning (*Momsoms*) för det j:te företaget i den i:te branschen vid tidpunkt t

X_{2ijt} socialavgiftsuppgifter (*Momssocbet*) för det j:te företaget i den i:te branschen vid tidpunkt t

X_{3ijt} Moms till godo (*Momstg*) för det j:te företaget i den i:te branschen vid tidpunkt t

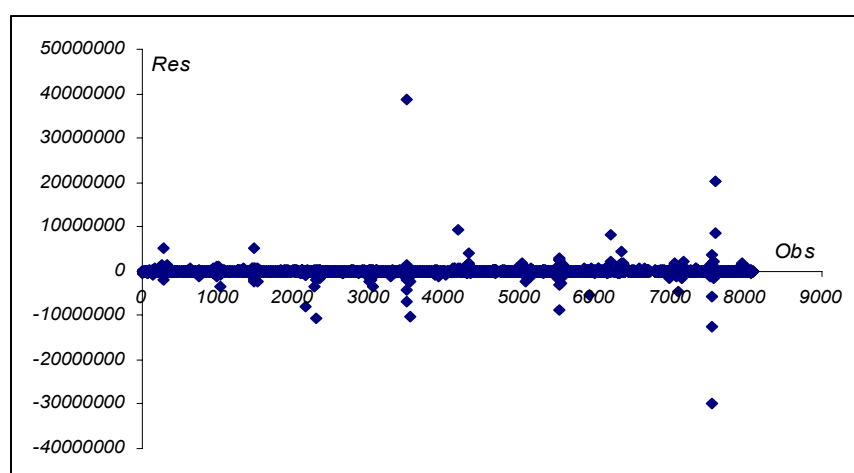
Modell 1 är en vanlig linjär modell som har anpassats i jämförelsesyfte. Som enda förklarande variabel ingår företagets omsättning, *Momsoms*. En linjär modell bygger på antagandet att slumpstörningarna ε_{ijt} är oberoende. Ett diagram över residualerna (ej redovisat) visar att de bildar systematiska mönster och att modell 1 därför inte är en lämplig modell. Modell 2 är en mixad linjär modell med slumpeffektkomponenter som hanterar korrelation mellan företag i samma bransch samt korrelation mellan varje observation av respektive företag. Endast *Momsoms* ingår som förklarande variabel. Jämfört med modell 1 är AIC-värdet lägre, och ett diagram över residualerna (ej redovisat) visar att en stor del av de systematiska avvikelserna kunnat förklaras. Modell 3 är en vidareutveckling av modell 2. Fortfarande ingår endast *Momsoms* som förklarande variabel. Här har skapats indikatorvariabler för respektive bransch, och dessa indikatorvariabler har sedan multiplicerats med modellens förklarande variabel. Syftet med detta är att bestämma en REM-modell som tillåter företagsspecifika skattningar av avvikelser från den gemensamma lutningsparametern. Företag 2021 har lämnats som referens. AIC-värdet sjönk anmärkningsvärt men samtidigt ökade modellens komplexitet från två fixa effekter i modell 1 och 2 till hela 2022 fixa effekter. I modell 4, 5 och 6 har övriga förklarande variabler som visats ha teoretisk och praktiskt förankring i modellen inkluderats. Modell 4 innehåller förutom *Momsoms* även företagets socialavgiftsuppgifter (*Momssocbet*) som förklarande variabler. Modell 5 innehåller dessutom uppgifter om företagets moms till godo (*Momstg*). I denna modell har emellertid branscherna beskrivits med fixa effekter genom att en indikatorvariabel har skapats för varje bransch. Modell 6 innehåller samma förklarande variabler som modell 5 men här har branscherna återigen beskrivits med slumpeffekter.

Med AIC-värdet som modellvalskriterium är modell tre den bästa modellen. Dock var inte en enda av de 2 020 skattade parametrarna för avvikelser från den gemensamma lutningsparametern signifikant skilda från noll. Parameterskattningarna för dessa avvikelser var mycket nära noll även i praktiken. Dessutom blev **G**-matrisen ej positivt definit. Att skattade kovariansmatriser ej blir positivt definita är ofta ett tecken på stor osäkerhet i kovariansskattningarna (*SAS Users Manual*, kapitel 18), i detta fall sannolikt till följd av att modellen överanpassats. Baserat på detta betraktades modellen trots sitt låga AIC-värde inte som en förbättring av den ursprungliga modellen. En enklare modell där företagets responsvariabel beskrivs med en gemensam lutningsparameter är därför att föredra.

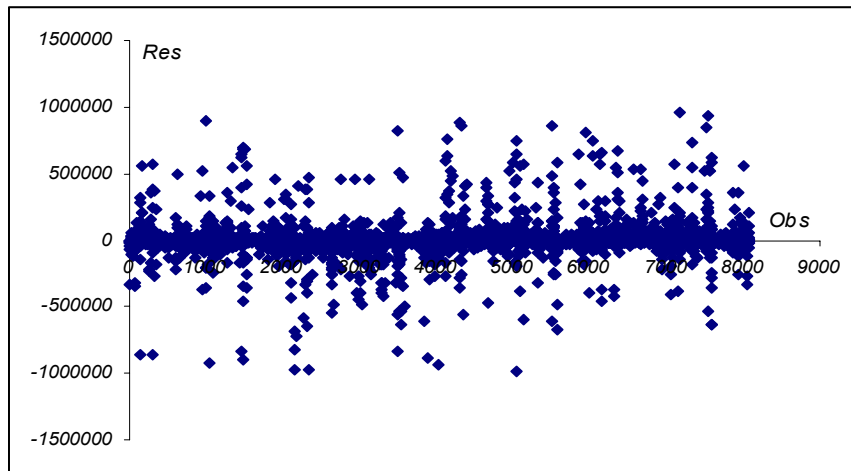
Modellerna med näst lägst AIC-värden var modell 5 och 6. Modell 5 har något lägre AIC-värde, men i valet mellan en modell med en kombination av fixa effekter och slumpeffekter eller en modell med enbart slumpeffekter är traditionen att välja modellen med enbart slumpeffekter (*Baltagi m.fl.*, 2001). För modell 6 gäller att samtliga parameterskattningar för de förklarande variablerna var signifikanta. När det gäller skattningarna av slumpeffekterna gäller det gemensamt för alla modeller där dessa inkluderats att de för många företag har höga värden på punktskattningen men samtidigt mycket hög standardavvikelse. Detta medför att slumpeffekterna ej blir signifikant skilda från noll på femprocentnivån. Därmed väcks förstås

frågan om det verkligen är nödvändigt att inkludera slump effekterna i modellen. I detta sammanhang är det dock viktigt att påminna sig vad slump effekterna har för syfte. Slump effekterna ger de branschspecifika respektive företags specifika avvikelserna från det gemensamma interceptet. Om ett företag har en ickesignifikant skattning av lutningsparametern innebär detta att dess skattning av den företags specifika avvikelsen från det gemensamma interceptet ej är signifikant skild från noll och därmed att företaget beskrivs bra av det gemensamma interceptet. Men vissa företag har faktiskt signifikanta slump effekter, och det är dessa företag som avviker från det gemensamma interceptet. Företagen med signifikanta slump effekter är de som är riktigt små samt, viktigare, de som är riktigt stora. Som det redan har konstaterats finns det ett tiotal mycket stora företag. Det går inte att enbart anpassa slump effekter för dessa företag, antingen måste parametrarna anpassas till alla företag eller också inte till något alls. Eftersom det finns företag som avviker kraftigt från övriga rättfärdigas därigenom slump effekternas syfte i modellen. Modell 6 undersöktes därför närmare. Figur 4.4 – 4.5 innehåller diagram över modellens residualer, sorterade efter år och bransch.

Fig 4.4 Residualerna för modell 6



I diagrammet framträder några observationer med mycket stora avvikelser. Dessa tillhör de största företagen. För att bättre åskådliggöra övriga residualer skapades ett filter som sorterade bort alla residualer med avvikelser större än $\pm 1\,000\,000$. Totalt 92 observationer filterades bort.

Fig 4.5 Residualerna för modell 6, filtrerade en gång

Residualerna verkar falla slumpmässigt. Samtliga parameterskattningar är signifikanta på fem procents signifikansnivå och residualerna uppvisar oberoende. Sett till de klassiska modellvalskriterierna är därmed modell 6 en god modell.

Viktigt att observera är att residualplottarna ovan avser residualerna på företagsnivå, alltså differensen mellan observerade och prognostiserade värden för respektive företag. Detta innebär att residualavvikelseerna på branschnivå och för hela det privata näringslivet till viss del kan komma att ta ut varandra i de slutliga prognoserna.

4.3 Parameterskattning

Den modell som valts som den bästa har alltså utseendet

$$Y_{ijt} = \alpha_0 + \alpha_1 X_{1ijt} + \alpha_2 X_{2ijt} + \alpha_3 X_{3ijt} + \mu_i + v_{ij} + \varepsilon_{ijt}$$

α_0 betecknar modellens intercept, α_1 är parameter för variabeln *Momsoms*, α_2 parameter för variabeln *Momssocbet* och α_3 parameter för variabeln *Momstg*. μ_i , v_{ij} och ε_{ijt} är modellens slump effekter och hanterar korrelation mellan företag i samma bransch samt korrelation mellan varje observation av respektive företag.

Modellexekveringen gav följande parameterskattningar. Kolumnen *Skattning* innehåller parametrarnas punktskattningar och *Stdavv* standardfelet. *t* är teststatistikan i det klassiska t-testet (kvoten mellan punktskattning och standardfel) och *Pr > p* anger sannolikheten för typ I-fel (risken att förkasta en sann nollhypotes). *Undre* och *Övre* utgör undre respektive övre gräns i ett 95-procentigt konfidensintervall för α -parametern.

Tabell 4.3 Parameterskattningar av fixa effekter

Fix effekt	Skattning	Stdavv	t	Pr > p	Undre	Övre
α_0	20189	22255	0,91	0,3908	-31132	71510
α_1	0,4812	0,04267	19,85	<0,0001	0,7633	0,9306
α_2	0,8470	0,003322	144,86	<0,0001	0,4747	0,4877
α_3	-2,2663	0,05871	-38,60	<0,0001	-2,3814	-2,1512

Tabell 4.4 Parameterskattningar av slump effekter

Slump-effekt	Bransch	Företag	Skattning	Stdavv	t-kvot	Pr > p
μ_1	1		1921,46	24619	0,08	0,9378
...
μ_9	9		-4349,45	22580	-0,19	0,8473
$v_{1,1}$	1	1	-14132	412095	-0,03	0,9726
...
$v_{9,2020}$	9	2020	13257	467622	0,03	0,9774

Tabell 4.5 Parameterskattningar av kovariansparametrar

Kovarians-parametrar	Skattning	Undre	Övre
σ_μ^2	6,6007E8	66081036	1,127E82
σ_v^2	5,176E11	4,72E11	5,7E11
σ_ε^2	7,554E11		

Det framgår ur tabell 4.3 att interceptskattningen ej är signifikant skild från noll. Detta är logiskt: om företaget inte har någon omsättning och inte har gjort några inköp har det per definition förädlingsvärde noll. Trots detta inkluderades interceptet i modellen. Som väntat fick momsvariabeln negativ parameterskattning. De fixa effekterna ska tolkas på följande sätt. Om företagets omsättning (*Momsoms*) ökar med en euro ökar förädlingsvärdet med 0,48 euro. Om företagets socialavgiftsinbetalningar (*Momssocbet*) ökar med en euro ökar förädlingsvärdet med 0,85 euro. Om den moms företaget har till godo från sina inköp (*Momstg*) ökar med en euro minskar förädlingsvärdet med 2,27 euro. Ur tabell 4.4 har parameterskattningarna för de branschspecifika slump effekterna μ_2, \dots, μ_8 samt parameterskattningarna för de företagsspecifika slump effekterna $v_{1,2}, \dots, v_{9,2019}$ av utrymmesskäl klippts bort. Parameterskattningen för μ_1 ska tolkas på följande sätt. Om en observation tillhör bransch 1 blir interceptskattningen $20189 + 1921,46 = 22110,5$. Parameterskattningen för $v_{1,1}$ ska tolkas som att om observationen tillhör företag 1 (och företag 1 tillhör bransch 1) blir interceptet $20189 + 1921,46 + (-14132) = 7978,46$. Övriga parameterskattningar tolkas analogt.

4.4 Modellens validitet

Modellen baserades på antagandet att slump effekterna skulle vara oberoende och normalfördelade. Oberoendet skulle gälla såväl inom som mellan de olika grupperna av slump effekter.

Oberoende för de tre typerna av slump effekter (μ_i , v_{ij} , ε_{ijt}) undersöktes genom att en korrelationsmatris beräknades (tabell 4.6).

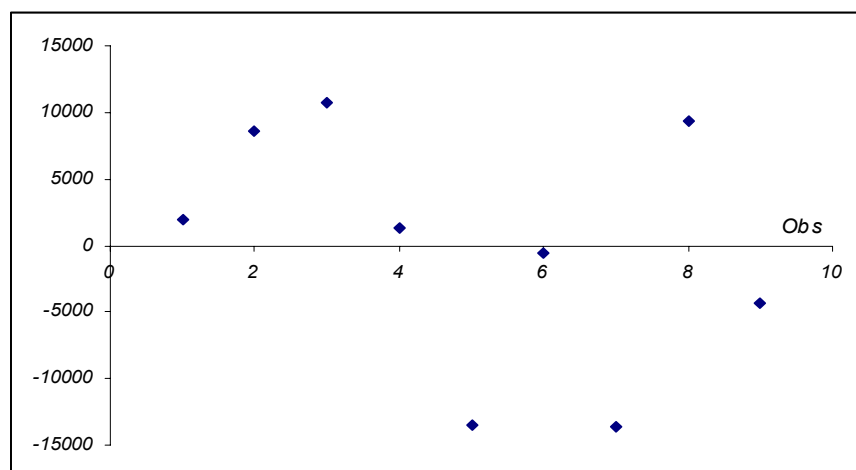
Tabell 4.6 Korrelationsmatris för slump effekter

	μ_i	v_{ij}	ε_{ijt}
μ_i	1	-0,216	-0,405
v_{ij}	-0,216	1	0,332
ε_{ijt}	-0,405	0,332	1

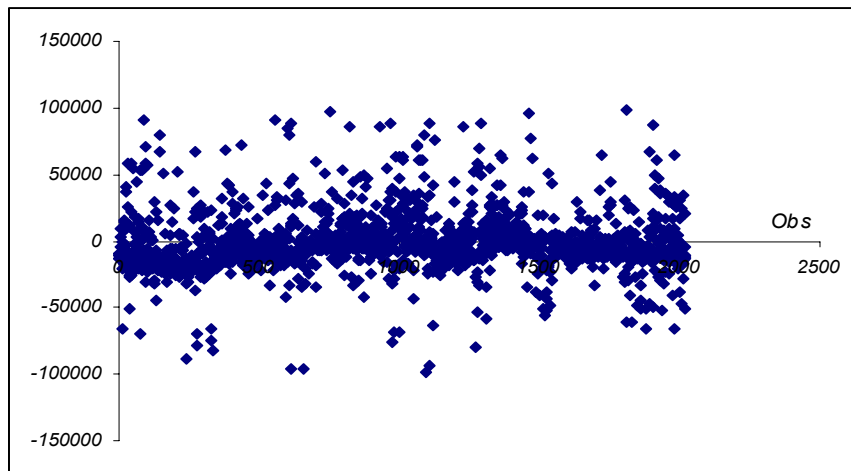
Tabell 4.6 visar att det finns en svagt negativ korrelation mellan slump effekterna för branscher och företag samt mellan branscher och den återstående slump störningen. Vidare finns det en svagt positiv korrelation mellan slump effekterna för företag och den återstående slump störningen. Korrelationerna är dock inte anmärkningsvärt höga och för att behålla en så enkel modell som möjligt bortses från dessa korrelationer.

Förekomst av outliers bland slump effekterna kontrollerades genom grafiska framställningar av modellens slump effekter. Outliers gör att en variabel ej kan betraktas som normalfördelad.

Fig 4.6 Slump effekterna μ_i



Figur 4.6 visar de branschspecifika slump effekterna μ_i plottat i branschernas ordning. Som framgår ur figuren finns inga outliers bland slump effekterna μ_i .

Fig 4.7 Slumpeffekterna v_{ij} 

Figur 4.7 visar de företagsspecifika slumpeffektkomponenterna v_{ij} plottat i företagens ordning. I syfte att förtydliga figuren tydligare observationer med avvikelser större än $\pm 100\,000$ euro, sammanlagt 141 slumpeffekter, filterats bort. Detta innebär att 141 av de 2 021 slumpeffekterna för företag utgör outliers.

Slumpeffekterna ε_{ijt} studerades i figur 4.4 – 4.6. Bland dessa utgör sammanlagt 592 av de 12 156 återstående slumpstörningarna outliers.

Av det stora antalet outliers hos slumpeffekterna framgår att antagandet om normalfördelning ej uppfylls. Detta har två konsekvenser. Dels kommer osäkerhetsintervall, exempelvis vid intervallskattningar, ej att vara meningsfulla. Det finns dock inga enkla lösningar på detta problem. Dessutom har den valda modellen visat sig kunna förklara en stor del av variationen i observerade data.

Slutsatsen blir alltså att den föreslagna modellen har ett praktiskt värde för att prognostisera förändringar i BNP, men osäkerhetsbedömningarna bör tas med viss reservation.

4.5 Felkällor i modellen

Mixade linjära modeller är känsliga för avvikande observationer. Två typer av avvikande observationer kan förekomma: avvikande observationer mellan företag samt avvikande observationer inom samma företag vid olika tidpunkter. Avvikande observationer mellan företag påverkar skattningen av \mathbf{G} medan avvikande observationer inom företag påverkar skattningen av \mathbf{R} . Det åländska näringslivet består av många jämnstora företag och ett litet antal mycket stora företag. De stora företagen har så pass stor påverkan på det åländska näringslivet att resultaten skulle förskjutas om man plockade bort dem ur analysen. Dessa utgör därför avvikande observationer mellan företag. Vissa av de största företagen ingår ej bland de största varje studerat år. Därigenom framkommer även avvikande observationer

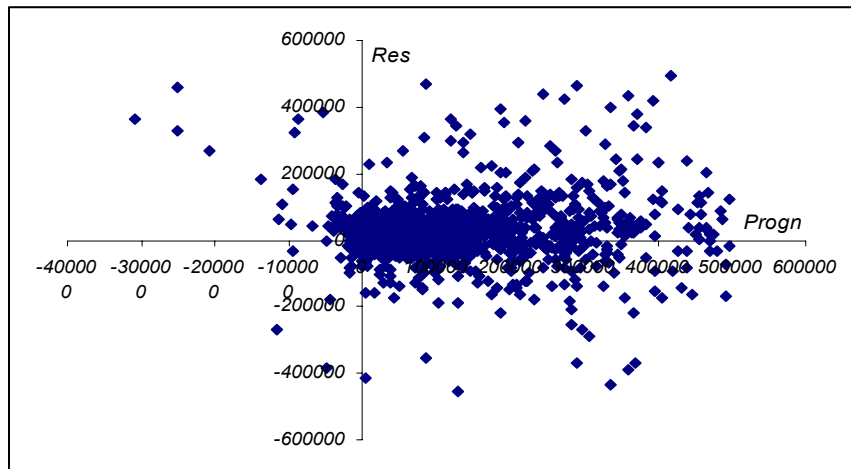
inom företag. Följaktligen finns risk för att modellens variansskattningar för slumpeffekterna (alltså skattningarna av kovariansparametrarna) i vissa fall är felaktiga. Kovariansparametrarna har ingen betydelse för modellens prognostiserande egenskaper men skulle kunna leda till felaktig specificering av modellen. Detta följer ur att parameterskattningarnas signifikans kontrolleras med t-värdet, och t-värdet beräknas som kvoten mellan parameterns punktskattning och dess standardavvikelse. Om standardavvikelseerna exempelvis systematiskt har överskattats skulle detta leda till för låga t-kvoter och att parametrar som egentligen är signifikanta ej skulle komma att inkluderas i modellen. Detta skulle kunna vara en anledning till att så pass få slumpeffekter faktiskt blev signifikanta. Emellertid inkluderades slumpeffekterna i modellen med teoretiska argument och därmed bortsågs från parameterskattningarnas signifikans. Detta innebär att avvikande värden inte utgör någon allvarlig felkälla.

Företagen får tillgodo den moms de betalar i samband med inköp och investeringar. Förädlingsvärde definieras som differensen mellan företagets omsättning och insatsförbrukning. I insatsförbrukningen ingår per definition företagets inköp men ej deras investeringar. Skillnader mellan vad som är inköp och vad som är investeringar framkommer ej i momsregistret och därmed ligger information om både företagets inköp och investeringar i variabeln *Momstg*. Detta innebär att det föreligger ett systematiskt fel i variabeln *Momstg*. Felets storlek varierar med vilken typ av företag det gäller. För ett kapitalintensivt företag blir felet av naturliga skäl större än för ett arbetskraftsintensivt företag. I samråd med uppdragsgivaren beslutades att acceptera denna felkälla.

Nedan diskuteras några generella felkällor i ekonometriska analyser enligt *Gujarati* (1995).

Multikollinearitet innebär att det finns ett linjärt samband mellan de förklarande variablerna i modellen. Ett linjärt samband innebär i praktiken att flera av modellens förklarande variabler innehåller samma information. I den för detta projekt valda modellen ingår tre förklarande variabler: *Momsoms*, *Momssocbet* samt *Momstg*. Vid val av modellens förklarande variabler har bland annat risken för multikollinearitet beaktats, men givetvis går det inte att utesluta viss risk för multikollinearitet i modellen. I tabell 4.1 framgick att det föreligger ganska hög korrelation mellan variablerna *Momsoms* och *Momssocbet*. Multikollinearitet skapar stor osäkerhet i parameterskattningarna. Följderna blir likartade de för avvikande värden men påverkar istället variansskattningarna för parametrarna för de fixa effekterna. De förklarande variablerna har dock framförallt valts med teoretiska argument, och därmed borde problemet till viss del undkommas.

Heteroscedasticitet innebär att slumpstörningarna ε_{ijt} har ickekonstant varians. Heteroscedasticitet upptäcks enklast i plottar av residualerna mot modellens prognostiserade värden. Om residualerna uppvisar tydliga mönster, exempelvis genom att vara trattformade, finns det anledning att misstänka heteroscedasticitet. Heteroscedasticitet resulterar liksom multikollinearitet i osäkerhet i parameterskattningarna.

Fig 4.8 Residualerna plottade mot prognostiserade värden

Figur 4.8 åskådliggör modellens residualer plottade mot dess prognostiserade värden. De 467 mest avvikande observationerna har sorterats bort. Ur figuren framgår inga tydliga mönster, men filtreringen innebär att man inte enbart baserat på figuren kan dra slutsatsen att det inte föreligger någon heteroscedasticitet. Liksom för multikollinearitet har dock konsekvenserna av eventuell förekomst av heteroscedasticitet mildrats genom att de förklarande variablerna framförallt har valts med teoretiska argument.

Autokorrelation innebär att det finns korrelation mellan slumpstörningarna ε_{ijt} . Även autokorrelation framgår i residualplottar. Autokorrelationen skulle skapa systematiska mönster bland residualerna. Bortsett från de allra största företagen, vars kraftiga avvikelser redan har konstaterats, går det ej att påvisa några tydliga mönster bland residualerna.

5. Tillämpning

I detta kapitel används den modell som anpassats till data för att beräkna de årliga procentuella förändringarna av det privata näringslivets tillskott till BNP för år 2000 och 2001. Eftersom bokslutsdata ej fanns tillgängligt för år 2000 och 2001 vid tidpunkten för detta projekts genomförande kan förfarandet alltså i viss bemärkelse liknas vid prognostisering.

Vid tidpunkten för detta projekts genomförande fanns fullständiga bokslutsuppgifter och momsregisteruppgifter för år 1996 – 1999. Dessa data användes i föregående kapitel för att bestämma modellens parametrar. Men momsregistret innehöll dessutom data för år 2000 och 2001. Genom att sätta in dessa data i regressionsfunktionen erhöles prognoser av förädlingsvärdet för år 2000 och 2001.

5.1 Branschvis

Genom att summera de prognostiserade förädlingsvärdena för varje företag erhöles branschvisa prognoser i absoluta tal av förädlingsvärdena. Observera dock att prognoserna i absoluta tal systematiskt underskattar det privata näringslivets tillskott till BNP eftersom inte alla företag ingår i analysen. Observera även, som diskuterades i kapitel 2.4, att antalet observationer i varje bransch i vissa fall är för få för att de branschvisa prognoserna ska vara helt tillförlitliga.

Tabell 5.1 Prognostiserade förädlingsvärden branschvis för år 2000 (euro)

Bransch	Prognostiserat förädlingsvärde
Primärnäringsar	5201248
Livsmedelsindustri	6960734
Övrig industri	36021762
Byggverksamhet	20238640
Handel	52440244
Hotell och restaurang	11632575
Transp. och kommunikation	308753930
Bank och försäkring	19843291
Övriga tjänster	19587593

Tabell 5.2 Prognostiserade förädlingsvärden branschvis för år 2001 (euro)

Bransch	Prognostiserat förädlingsvärde
Primärnärings	5552799
Livsmedelsindustri	9148928
Övrig industri	40549572
Byggverksamhet	20026402
Handel	49765677
Hotell och restaurang	10931689
Transp. och kommunikation	318901491
Bank och försäkring	20107215
Övriga tjänster	21595898

I tabell 5.1 och 5.2 redovisas de prognostiserade förädlingsvärdena för respektive bransch år 2000 och 2001. Prognoserna indikerar ökningarna i alla branscher utom *Byggverksamhet*, *Handel* samt *Hotell och restaurang*.

I figur 5.1 – 5.9 har de årliga procentuella förändringarna hos observerade respektive prognostiserade förädlingsvärden plottats branschvis. Förändringarna har beräknats som kvoten mellan de summerade förädlingsvärdena för ett år i en viss bransch dividerat med de summerade förädlingsvärdena i samma bransch föregående år. Vid summeringarna har endast de företag som existerar två på varandra följande år inkluderats. Syftet med detta var att minska inverkan från saknade observationer. Varje punkt i diagrammen utgör alltså en kvot mellan ett års summerade förädlingsvärde och det summerade förädlingsvärdet för samma grupp företag föregående år, multiplicerat med 100. Om punkten ligger exakt på y-axelns nivå 100 procent innebär detta att den procentuella förändringen av branschens förädlingsvärde sedan föregående år är noll.

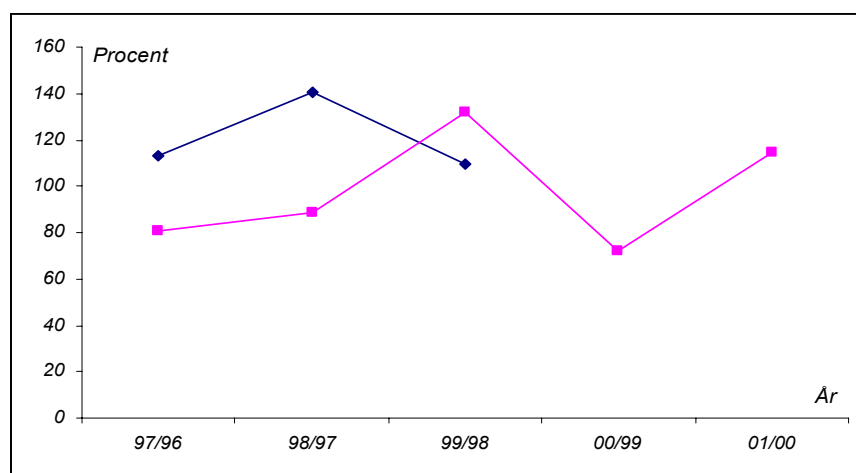
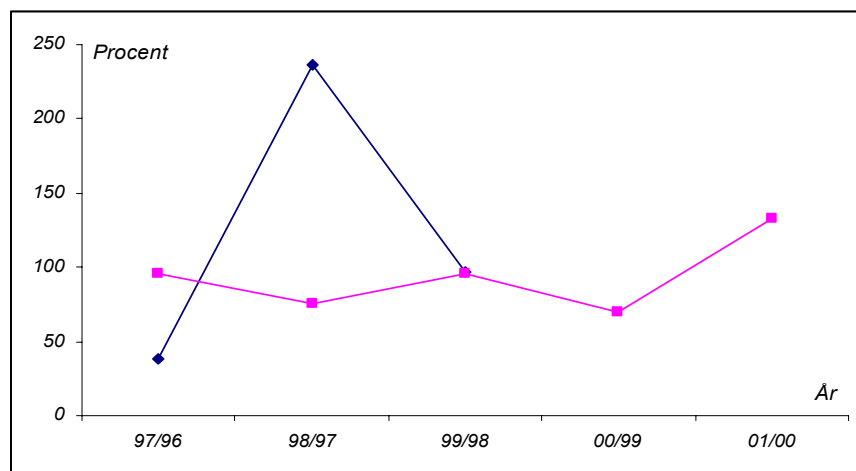
Fig 5.1 Årliga procentuella förändringar av förädlingsvärdet i branschen Primärnärings

Fig 5.2 Årliga procentuella förändringar av förädlingsvärdet i branschen Livsmedelsindustri



För branschen *Primärnäringsar* har observerats årliga ökningar om mellan fem och 40 procentenheter under perioden. De prognostiserade förändringarna indikerar minskningar utom för år 1999 och 2001. Sett till prognostiseringsfelen finns det anledning att misstänka att prognoserna för år 2000 och 2001 är överdrivet pessimistiska. Branschen *Livsmedelsindustri* hade en mycket djup svacka år 1997, vilket i diagrammet framgår som en kraftig procentuell ökning 1998. 1999 låg förädlingsvärdena kvar på 1998 års nivå. De prognostiserade förändringarna följer en betydligt jämnare trend och indikerar en minskning med några procentenheter år 2000 följt av en ökning i samma storleksordning år 2001.

Fig 5.3 Årliga procentuella förändringar av förädlingsvärdet i branschen Övrig industri

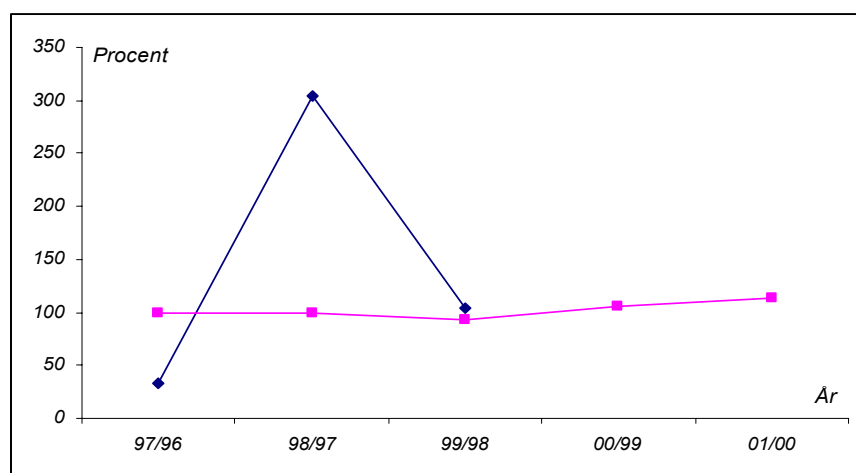
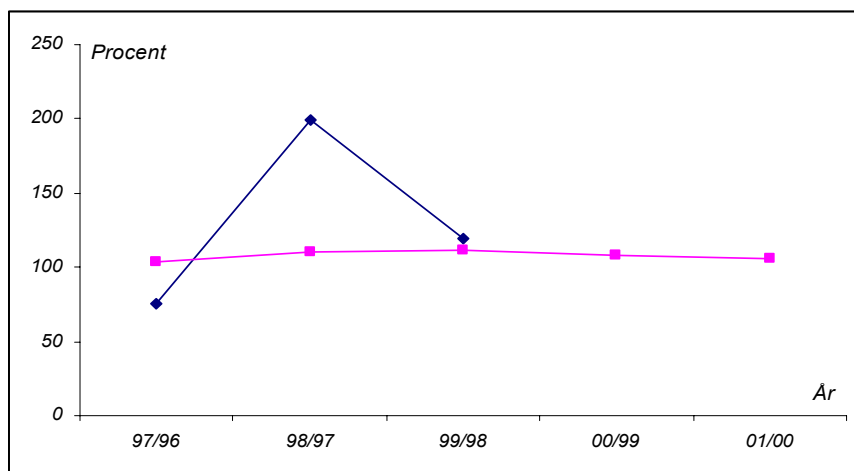


Fig 5.4 Årliga procentuella förändringar av förädlingsvärdet i branschen Byggverksamhet



Branschen *Övrig industri* gjorde en djupdykning 1997, vilket framgår som en kraftig procentuell ökning 1998. 1999 låg branschens summerade förädlingsvärden kvar på 1998 års nivå. De prognostiserade förändringarna har en betydligt jämnare karaktär än de observerade. Prognosen för år 2000 och 2001 pekar på svaga uppgångar bägge åren. Även branschen *Byggverksamhet* hade en svacka år 1997. De prognostiserade förändringarna följer en jämn trend med årliga ökningarna om någon eller några procentenheter.

Fig 5.5 Årliga procentuella förändringar av förädlingsvärdet i branschen Handel

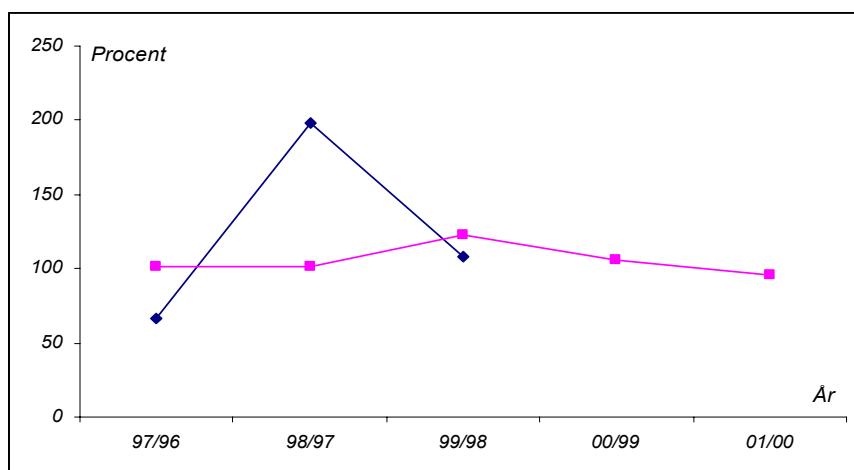
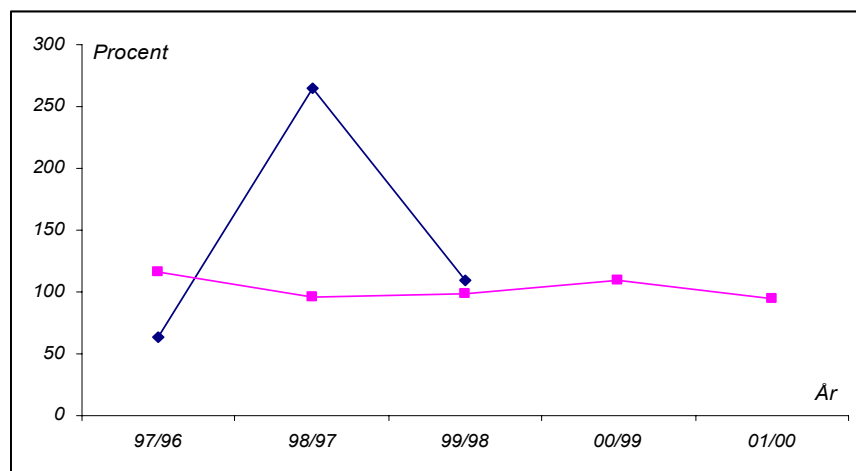


Fig 5.6 Årliga procentuella förändringar av förädlingsvärdet i branschen Hotell och restaurang



Bortsett från svackan år 1997 har de årliga prognostiserade förändringarna för branschen *Handel* god överensstämmelse med de observerade. Prognosen indikerar att branschens summerade förädlingsvärde år 2000 ligger kvar på 1999 års nivå och att branschen får en svag nedgång år 2001. Branschen *Hotell och restaurang* hade en kraftig svacka år 1997 men återhämtade sig sedan snabbt. 1999 gjorde branschen en viss ökning. De prognostiserade förändringarna följer de observerade värdenas medelvärde och pekar på en jämn trend för år 2000 och 2001.

Fig 5.7 Årliga procentuella förändringar av förädlingsvärdet i branschen Transport och kommunikation

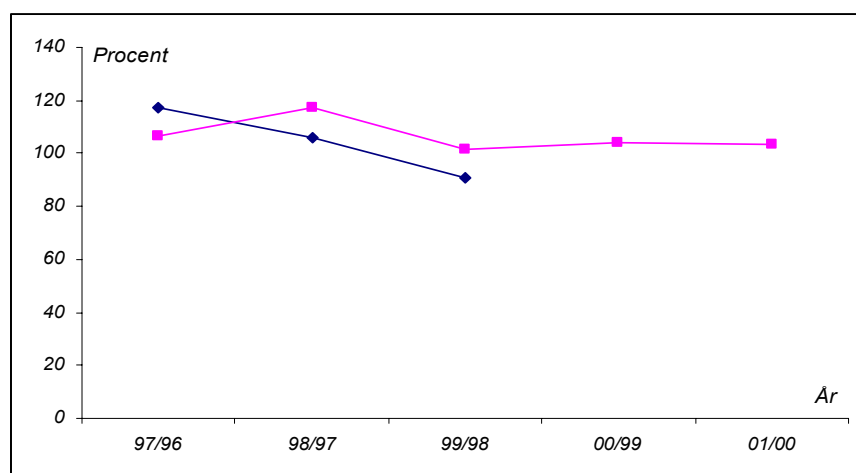
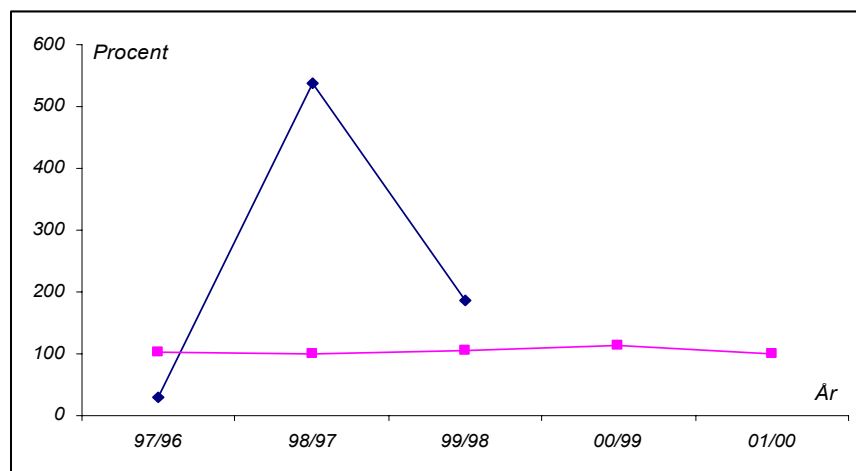
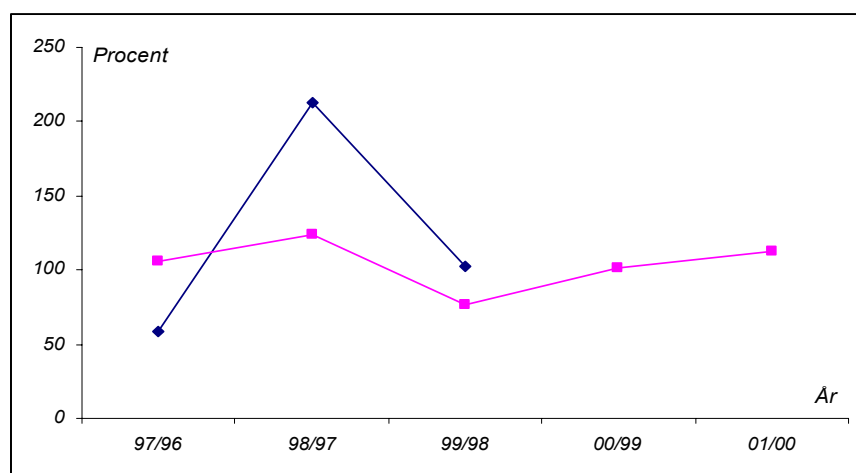


Fig 5.8 Årliga procentuella förändringar av förädlingsvärdet i branschen Bank och försäkring



Branschen *Transport och kommunikation* uppvisade en ökning om 20 procentenheter 1997 följt av en mindre ökning 1998 och en svag nedgång 1999. De prognostiserade förändringarna ligger nära de observerade och pekar på svaga uppgångar både år 2000 och 2001. Prognoseerna har goda förutsättningar att vara rättvisande. Branschen *Bank och försäkring* genomgick en mycket kraftig svacka 1997 som dock åtföljdes av en ordentlig uppgång 1999. Prognosen ligger i underkant av de observerade förändringarna. Prognosen för år 2000 och 2001 kan därför förväntas vara underskattande.

Fig 5.9 Årliga procentuella förändringar av förädlingsvärdet i branschen Övriga tjänster



För branschen *Övriga tjänster* har de prognostiserade förändringarna, bortsett från år 1998, god överensstämmelse med de observerade. Detta talar för en god prognos av 2000 och 2001 års procentuella förändringar av branschens summerade förädlingsvärden.

5.2 Hela det privata näringslivet

De prognostiserade förädlingsvärdena för branscherna summerades årsvis. Tabell 5.3 visar de prognostiserade förädlingsvärdena för år 2000 och 2001 i euro. På grund av att inte alla det privata näringslivets företag ingår i analysen ger de prognostiserade förädlingsvärdena i absoluta tal underskattningar av det privata näringslivets faktiska tillskott till BNP.

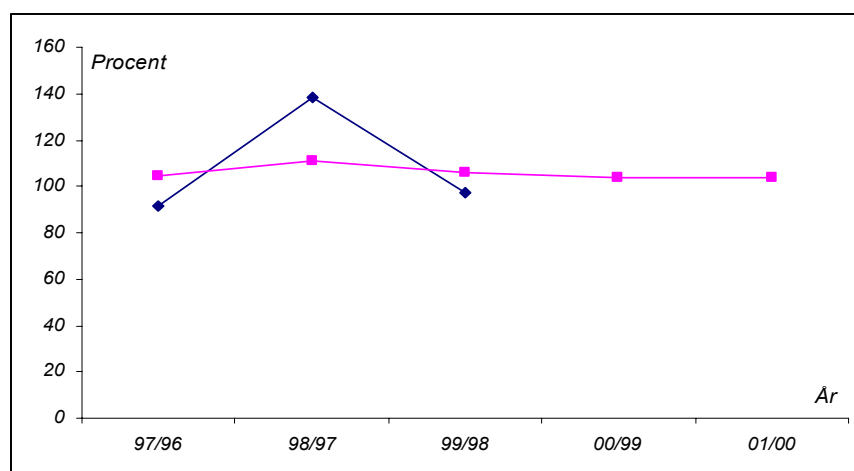
Tabell 5.3 Prognostiserade förädlingsvärden årsvis (euro)

År	Prognostiserat förädlingsvärde
2000	480680017
2001	496579671

Förutsatt att det förhållande som hittills gällt mellan förädlingsvärde och modellens förklarande variabler fortsätter gälla år 2000 och 2001, kommer det privata näringslivets tillskott till BNP år 2000 att vara 480 680 017 euro och år 2001 496 579 671 euro. Detta innebär en svag uppgång i det privata näringslivets tillskott till BNP år 2000 och 2001.

Figur 5.10 åskådliggör de årliga procentuella förändringarna hos det privata näringslivets tillskott till BNP. De årliga procentuella förändringarna har beräknats som kvoten mellan det privata näringslivets tillskott till BNP ett visst år dividerat med det privata näringslivets tillskott till BNP för samma grupp företag föregående år, multiplicerat med 100. Liksom för de branschvisa prognoserna ingår endast de företag som existerat två på varandra följande år. Syftet med detta är att minska problemen med saknade observationer. Diagrammet ska tolkas som att om en punkt ligger exakt på y-axelns nivå 100 procent innebär detta att den procentuella förändringen från föregående år är noll.

Fig 5.10 Observerade och prognostiserade årliga förändringar av det privata näringslivets förädlingsvärde



Till följd av nedgången i data år 1997 blev den procentuella förändringen 1998 mycket kraftig. De prognostiserade årliga förändringarna håller en jämn trend med årliga ökningarna om

några procentenheter. För år 2000 förväntas en ökning med 3,7 procentenheter jämfört med de prognostiserade värdena år 1999 eller en ökning med 6,5 procentenheter jämfört med de observerade värdena år 1999. För år 2001 förväntas en ökning med 3,9 procentenheter jämfört med de prognostiserade värdena år 2000.

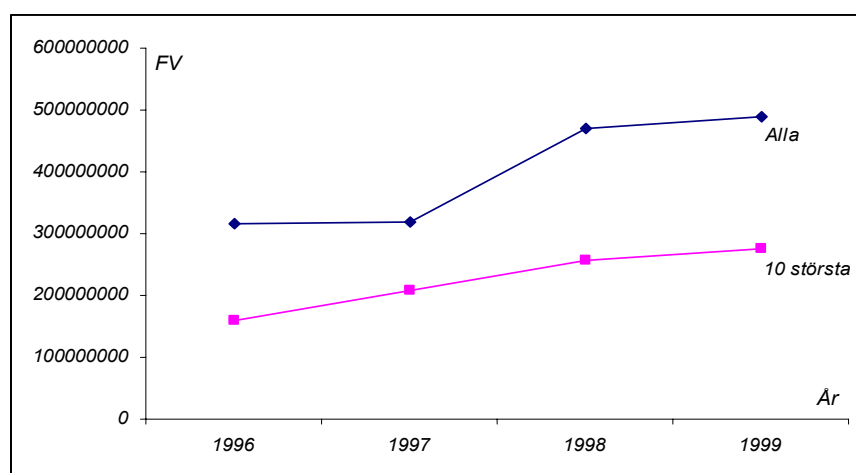
6. Reducerad metod

I detta kapitel genomförs en analys av data från de tio förädlingsvärdesmässigt största företagen. Syftet är att undersöka om en prognos som enbart baserats på dessa tio företag överensstämmer med prognosen som baserats på alla företag. Detta skulle ge en snabb och billig metod för att prognostisera de årliga procentuella förändringarna av det privata näringslivets tillskott till BNP.

6.1 Prognos

Sammanlagt tio företag och 25 observationer ingår i analysen. Sex av de tio företagen tillhör branschen *Transport och kommunikation*, två tillhör branschen *Bank och försäkring* och ett vardera tillhör branscherna *Livsmedelsindustri* samt *Övrig industri*. Fem av företagens observationer ingår i analysen vid alla de observerade tidpunkterna 1996 – 1999, två av dem ingår vid tre tidpunkter, två av dem vid två tidpunkter och ett vid endast en tidpunkt. För tre av företagen fanns observationer bland de kraftigt avvikande alla fyra åren, för två av dem tre år samt två stycken två år. Tre av företagen hade bara kraftigt avvikande värden ett år. Detta signalerar om fel i datamaterialet: det är inte sannolikt att ett företag blir så stort på bara ett år.

Fig 6.1 Alla företags summerade förädlingsvärden plottat mot det summerade förädlingsvärdet för de tio största företagen (euro)



Figur 6.1 åskådliggör storleksförhållandet avseende det årliga summerade förädlingsvärdet för de tio största företagen jämfört med alla företag (där även de tio största ingår). Som framgår ur figuren står de tio största företagen för omkring hälften av det årliga summerade förädlingsvärdet.

Första steget bestod i att anpassa en regressionsmodell till data. Sett till det låga antalet observationer valdes en enkel linjär regressionsmodell. Därmed bortsågs från paneldatastrukturen och därför även från eventuell korrelation mellan företag i samma

bransch och korrelation mellan varje observation av respektive företag.

Det resonemang kring vilka förklarande variabler som bör ingå i modellen som fördes för den fulla modellen kan även appliceras här. Parameterskattningen för variabeln *Momssocbet* blev dock kraftigt ickesignifikant och utelämnades därför ur modellen (det går dock inte att utesluta att parameterskattningens låga signifikans var en följd av multikollinearitet). Den bästa modellen fick följande utseende.

$$Y_{ijt} = \alpha_0 + \alpha_1 X_{1ijt} + \alpha_2 X_{3ijt} + \varepsilon_{ijt}$$

α_0 är modellens intercept, α_1 parameter för variabeln *Momsoms* och α_2 parameter för variabeln *Momstg*. ε_{ijt} utgör den återstående företagspecifika slumpstörningen. X_{1ijt} betecknar liksom för denna fulla modellen omsättning och X_{3ijt} moms till godo.

Tabell 6.1 Parameterskattningar av fixa effekter

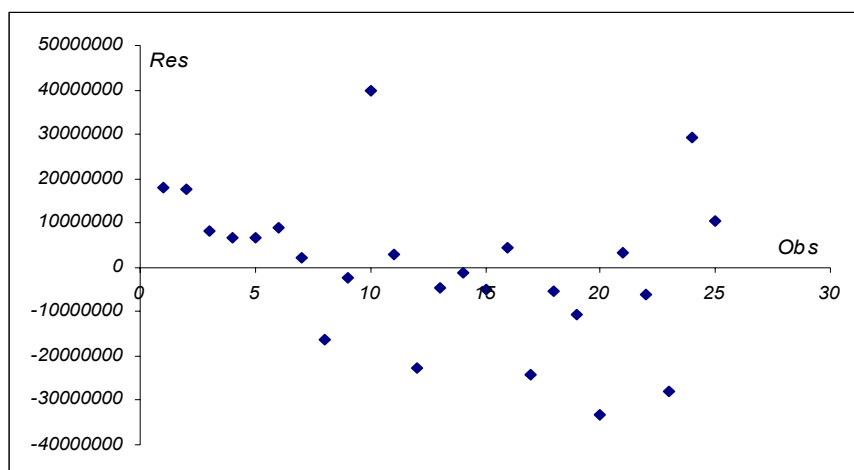
Fix effekt	Skattning	Stdavv	t	Pr > p	Undre	Övre
α_0	443262	5230515	0,08	0,9332	-1,04E7	11290685
α_1	0,5572	0,03212	17,35	<0,0001	0,4906	0,6238
α_2	-2,7040	1,2489	-2,17	0,0415	-5,2940	-0,1141

Tabell 6.2 Variansskattning för den återstående slumpstörningen

Var-skattning	Skattning
σ^2_ε	3,16E14

Den gemensamma interceptskattningen var 443 262 euro. Parametrarna ska tolkas som att om företagets omsättning ökar med en euro ökar förädlingsvärdet med 0,56 euro och om företagets moms till godo från alla inköp ökar med en euro minskar förädlingsvärdet med 2,7 euro.

Fig 6.2 Residualerna för den reducerade modellen



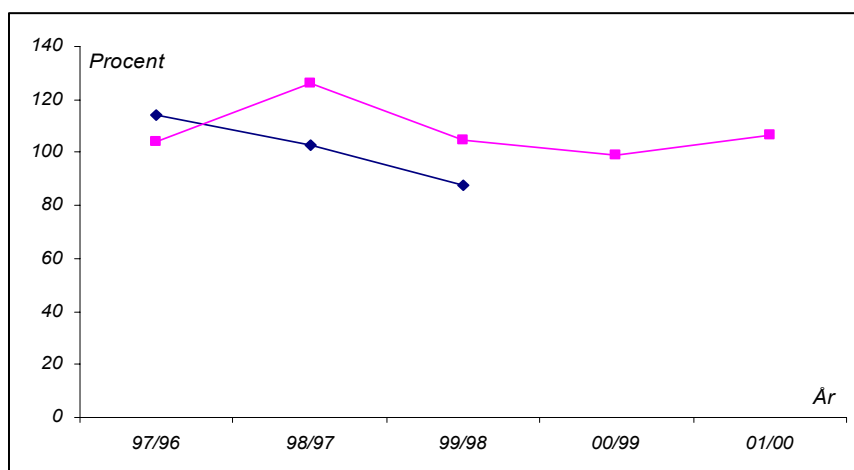
Figur 6.2 visar residualerna plottade i observationsordningen, sorterad först efter bransch och sedan efter år. Residualerna faller acceptabelt slumpmässigt, sett till det låga antalet observationer.

Tabell 6.3 Prognostiserade förädlingsvärden årsvis (euro)

År	Prognostiserat förädlingsvärde
2000	288105255
2001	307624227

Tabell 6.3 visar de tio största företagens prognostiserade förädlingsvärden för år 2000 och 2001 i absoluta tal. Figur 6.3 visar observerade samt prognostiserade årliga procentuella förändringar för de tio största företagen. Endast de företag som existerar två på varandra följande år har inkluderats. Om en punkt ligger exakt på y-axelns nivå 100 procent innebär detta att förändringen från föregående år är noll.

Fig 6.3 Observerade och prognostiserade förädlingsvärden för de tio största företagen



Ur figur 6.3 framgår att de prognostiserade värdena alla år utom 1997 överskattar de observerade. För år 2000 förväntas att det privata näringslivets tillskott till BNP ligger kvar på 1999 års nivå jämfört med de prognostiserade värdena 1999, möjligtvis med en viss ökning. Jämfört med de observerade värdena år 1999 förväntas en ökning med några procentenheter för år 2000. För år 2001 förväntas en ökning med 6,8 procentenheter.

6.2 Jämförelse

Sett till de prognostiserade procentuella förändringarna (jämför figur 5.10 och figur 6.3) framgår att de tio största företagen ensamma har mycket stor inverkan på det privata näringslivets tillskott till BNP. Prognosen för år 2000 och 2001 baserat på de tio största företagen blev något mer optimistisk än den som baserades på alla företag, men skillnaden är ej särskilt stor.

Sett till hanterligheten tack vare det låga antalet observationer samt den låga kostnaden jämfört med att studera alla företag utgör därför de tio största företagen en mycket god indikator av det privata näringslivets tillskott till BNP. Dock kan en prognos som endast baserats på de tio största företagen ej betraktas som representativ för hela det privata näringslivet på Åland. Detta följer ur att hela sex av de tio företagen tillhör branschen *Transport och kommunikation*, och därmed skulle förändringar som enbart påverkar denna bransch leda till felaktiga prognoser av de årliga förändringarna hos hela det privata näringslivet.

7. Slutdiskussion

Syftet med detta projektarbete var att bestämma en metod för att beskriva de åländska företagens förädlingsvärde som en funktion av de uppgifter som framkommer ur Skattestyrelsens månatliga momsregister. Metoden ska användas för att prognostisera det privata näringslivets tillskott till BNP innan de exakta uppgifterna om företagens förädlingsvärden som framkommer ur Skattestyrelsens bokslutsregister är kända.

Genom teoretiska resonemang och praktisk prövning kunde det visas att de data som framkommer ur moms- och bokslutsregistret bäst beskrivs med en mixad linjär modell. Detta följer ur att data är av typen obalanserad paneldata och att det finns korrelationer i datamaterialet. Den modell som slutgiltigt bestämdes som den bästa har utseendet

$$Y_{ijt} = \alpha_0 + \alpha_1 X_{1ijt} + \alpha_2 X_{2ijt} + \alpha_3 X_{3ijt} + \mu_i + v_{ij} + \varepsilon_{ijt}$$

där X_{1ijt} betecknar omsättning, X_{2ijt} betecknar socialavgiftsinbetalningar och X_{3ijt} betecknar moms till godo från alla inköp för det j :te företaget i den i :te branschen vid tidpunkt t . μ_i , v_{ij} och ε_{ijt} är slump effekter som tar hänsyn till korrelation mellan företag i samma bransch samt korrelation mellan varje tidsmässigt åtskild observation av respektive företag. Till följd av flera faktorer, bland annat företagsstrukturen på Åland, är data ej normalfördelat. Lämplig skattningsmetod är därför momentskattningsmetoden MINQUE.

Förutsatt att det finns data i momsregistret som är av senare datum än data i bokslutsregistret kan modellen användas för prognostisering. Följande steg krävs för att göra en prognos.

1. De nya uppgifterna från momsregistret och bokslutsregistret läggs till i det datamaterial som ställts samman för detta projekt. Förädlingsvärdena markeras som saknade observationer för de tidpunkter som ska prognostiseras.
2. Data läses in i ett statistikprogram och modellen exekveras. Om enbart momsregistret har uppdaterats kommer exekveringen att leda till prognostiserade förädlingsvärden för varje företag till och med den sista observationen från momsregistret. Om även bokslutsregistret har uppdaterats kommer exekveringen dessutom att resultera i nya skattningar av modellens parametrar. Detta innebär att modellens parametrar med tiden kommer att baseras på mer data och därmed förbättras. Tack vare detta kommer även prognosernas kvalitet att förbättras.
3. Genom att summera företagens prognostiserade förädlingsvärden över tidsperioden av intresse fås prognoser av det privata näringslivets tillskott till BNP. Dock kommer sådana summeringar, på grund av att det är svårt att inkludera alla företag i analysen, att leda till systematiska underskattningar av det privata näringslivets tillskott till BNP. Genom att istället studera de *årliga procentuella förändringarna* av det privata näringslivets tillskott till BNP undkommes problemet. Genom att summera förädlingsvärdet för de företag som ingår i datamaterialet två på varandra följande år och dividera med det summerade

förädlingsvärdet för samma grupp företag föregående år erhålls årliga procentuella förändringar som dessutom tar hänsyn till den påverkan på resultaten som företag som ingår det ena men inte det andra året kan ha på resultaten.

Under arbetets gång kunde det konstateras att tio företag hade förädlingsvärden som var mångdubbelt större än förädlingsvärdena för övriga företag. Hypotesprövning visade att en modell som enbart baserats på dessa tio företag utgör en god indikator av det privata näringslivets tillskott till BNP. Genom att enbart studera dessa företag kan man därför på ett snabbt och billigt sätt prognostisera det privata näringslivets tillskott till BNP. Dock är det viktigt att observera att dessa tio företag ej på ett generellt sätt i framtiden kan betraktas som representativa för hela det privata näringslivet på Åland, även om så var fallet för de data som fanns tillgängliga vid tidpunkten för detta projekts genomförande.

Modellerna tillämpades för att beräkna de årliga procentuella förändringarna av det privata näringslivets tillskott till BNP år 2000 och 2001. Såväl modellen som baserats på alla företag som modellen som enbart baserats på de tio största företagen indikerade att det privata näringslivets tillskott till BNP kommer att öka med några procentenheter både år 2000 och 2001.

Referenslista

Baltagi, B H (1999), *Econometric Analysis of Panel Data*, andra upplagan. Wiley & Sons, Chichester. ISBN 0-471-49937-4.

Baltagi, B H; Song, S H; Jung, B C (2001), ”The unbalanced nested error component regression model”. *Journal of Econometrics* **101**: 357 – 381.

Gujarati, D N (1995), *Basic Econometrics*. McGraw-Hill, Singapore. ISBN 0-07-113964-8.

Holmertz, G (1997), *Länder i fickformat: Finland*. Utrikespolitiska institutet, Stockholm. ISBN 91-7183-181-9.

Körner, S (1987), *Statistisk dataanalys*, andra upplagan. Studentlitteratur, Lund. ISBN 91-44-20522-8.

Lindström, B; Kinnunen, J (2003), *Det åländska jordbrukets framtida utvecklingsförutsättningar*, ÅSUB Rapport 2003:2.

Makridakis, S; Wheelwright, S C; Hyndman, R J (1998), *Forecasting – Methods and Applications*. John Wiley&Sons, New York. ISBN 0-471-53233-9.

Milliken, G A; Johnson D E (2002), *Analysis of Messy Data*, volym 3. CRC Press.

Nationalräkenskapsdata för Åland 1993 – 1996. Ålands statistik- och utredningsbyrå, Mariehamn. ISSN 1455-1985.

The MIXED Procedure, *SAS Users Manual*, kapitel 18. SAS Institute Inc.

Wu, C; Gumpertz, M L; Boos, D D (2001), ”Comparison of GEE, MINQUE, ML and REML estimating equations for normally distributed data”. *The American Statistician* **55**: 125 – 130.

Åland kort och gott (2002). Ålands landskapsstyrelse, Mariehamn. ISBN 951-8946-00-0.

Bilagor

Bilaga I

Nr	Beteckning	Förklaring
<i>Från Statistikcentralens företagsregister</i>		
1	Företag	Unikt värde (bara för detta datamaterial) 1-2021 för varje företag
2	Brn5	Femsiffrig branschkod med 349 klasser
3	Förklaring	Kräver ingen förklaring
4	Brn2	Tvåsiffrig branschkod med åtta klasser
5	Bransch	Branschkod med tio klasser
6	Branschnamn	Branschnamn för branschkod med tio klasser
7	Kommun	Kommunkod
8	År	År 1996 – 2001
<i>Från Skattestyrelsens bokslutsregister</i>		
9	Skattsoc	Socialutgifter
10	Skattsubbr	Driftsbidrag brutto
11	Skattsubne	Driftsbidrag netto
12	Skattsalary	Företagets löner och arvoden
13	Skattinput	Insatsförbrukningens värde
14	Skattoms	Företagets omsättning
15	SkattFV	Företagets förädlingsvärde, definieras som Skattoms – Skattinput
16	Skattlager	Lagerförändring, positiv om lagret ökat
<i>Från Skattestyrelsens momsregister</i>		
17	Momssalary	Alla löner och arvoden
18	Momsoms	Omsättning enligt bokföringen
19	Momssoc	Arbetsgivarens socialskyddsavgifter
20	Momssocjust	Minskning av socialskyddsavgift: justering för för hög tidigare inbetalning
21	Momssocbet	Arbetsgivarens slutliga socialskyddsavgifter
22	MomsEUsale	Försäljning av varor (ej tjänster) till andra EU-länder
23	MomsEUbuy	Inköp av varor (ej tjänster) från andra EU-länder
24	Momstg	Moms till godo från alla inköp (även investeringar)
25	Momsskabet	Moms som ska betalas in för företagets inhemska försäljning
26	Momsbet	Belopp som moms skyldigt företag faktiskt betalat in

Bilaga II

```
proc mixed method = mivque0;  
class Bransch Företag;  
model SkattFV = Momsoms Momssocbet Momstg;  
random Bransch Företag;  
run;
```

Kommandot *proc mixed* anropar den SAS-procedur som ska användas. Som skattningsmetod anges *mivque0*, vilket är SAS beteckning på MINQUE. Kommandot *class* anger vilka variabler som det ska skapas indikatorvariabler för och *model* anger förstås modellen. Kommandot *random* definierar modellens slump effekter, i det här fallet *Bransch* vilket ger en slump effekt för alla observationer som tillhör samma bransch samt *Företag* som ger en slump effekt för alla observationer som tillhör samma företag.