

Magisteruppsats

Department of Economics

Lund University

P.O. Box 7082

SE-220 07 Lund

SWEDEN

Författare: Nikolaos Alexandris och Måns Näsman

Titel: Prognostisering av småbolagsindex

Handledare: Thomas Elger och Birger Nilsson

Abstract: Analysarbetet av finansiella data i samband med prognostisering blir snabbt omfattande, både ur ett teoretiskt och ur ett empiriskt perspektiv. Prognosteorier står mot varandra. Är marknaden starkt effektiv eller finns det prognosmöjligheter? Att genomföra en empirisk prognosstudie resulterar snabbt i krävande beräkningsintensiva prognosmodeller. Då prognoser för flera tidpunkter behövs så måste beräkningarna upprepas och använda olika delmängder av tidsserierna som beräkningsunderlag. Transaktionskostnader uppstår i samband med handel och nedlagt arbetstid vid informationshämtning och analys. Om transaktionskostnaderna blir för stora försvinner möjligheten att utnyttja en eventuell prognosstyrka. En starkt effektiv finansiell marknad förhindrar prognosmöjligheter eftersom all tillgänglig information avspeglas i de aktuella aktiekurserna. Vi ifrågasätter styrkan i denna effektivitet. Ny information kan nå olika segment av marknaden olika snabbt. Prisjusteringar sker då gradvis. Internationella studier, t ex Arbeter (2003) och Lo & MacKinlay (1990), lyfter fram size-komponenten som en förklarande variabel. Vi undersöker motsvarande prognosmöjligheter på den svenska finansiella marknaden. Big Cap är stora bolag med hög omsättning och Small Cap är små bolag. Vår hypotes är att ny information prisjusterar Big Cap snabbare än Small Cap. Vi använder initialt en bivariat VAR-modell för att fånga upp eventuell korsautokorrelation mellan aktieslagen. Tidsserierna visade sig vara icke-stationära och uppvisade en långsiktig gemensam trend. Vi kompletterade därför metoden med en bivariat modell med felkorrigeringsterm. Engel och Grangers resultat inom prognostisering, felkorrigering och kointegration användes vid utvecklingen av denna andra prognosmodell. Som referensmodell valde vi en Random Walk prognosmodell. Modellerna utvärderades med hjälp av prognosutvärderingskriterierna ME, MAE och RMSE. Prognosutvärderingskriterierna visade på prognoskraft hos de bivariata prognosmodellerna eftersom de indikerade en bättre fit än för Random Walk prognoserna. Den ekonomiska signifikansen undersökte vi genom att göra en simulering som vi byggde upp med handelsstrategier baserade på våra tre prognosmodeller. Den ekonomiska signifikansen var starkare för de bivariata modellerna än för en passiv buy-and-hold strategi i Small Cap. Ett rullande estimeringsfönster på 500 värden användes för att ta fram prognoser. Då vårt argument bakom valet av storleken på prognosfönstret var för svagt så uppgraderade vi vår metod med möjligheten att variera prognosfönstrets storlek. Resultatet tillförde en dimension till vårt arbete, men oavsett storlek på fönstret så blev den huvudsakliga slutsatsen densamma. Size-komponenten har prognoskraft på den svenska marknaden och det kan finnas en gradvis informationsspridning från svenska Big Cap till Small Cap.

INNEHÅLLSFÖRTECKNING

1. INLEDNING.....	3
2. DATA & DATASTRÖMMAR.....	8
3. PROGNOSEMODELL.....	11
3.1 Prognosteorier	12
3.2 Setup av prognosmetoder.....	14
3.3 Val av prognosmodeller.....	15
3.4 Prognosmodell 1 - VAR.....	16
3.5 Långsiktiga samband	16
3.6 Prognosmodell 2 - VECM.....	18
3.7 Referensprognosmodell – Random Walk	20
3.8 Val av transformation	20
3.9 Utvärdering 1 - Prognosutvärderingskriterier	22
3.10 Utvärdering 2 – Simulering av handelsstrategier	23
3.11 Övergripande bedömningskriterier.....	23
4. RESULTAT	24
4.1 Inledning	24
4.2 Grafer över prognoser i jämförelse med verkligt utfall	25
4.3 Utvärdering 1 - Prognosutvärderingskriterier	28
4.4 Utvärdering 2 – Simulering av handelsstrategier	30
5. SLUTSATSER.....	32
6. BILAGOR.....	33
7. ORDLISTA.....	33
8. REFERENSER.....	34
BILAGA 1. Engel-Grangers 3-stegsmetod (Enders version).....	36
BILAGA 2. Programscript för prognosmodell 1 - VAR.....	38
BILAGA 3. Programscript för prognosmodell 2 - VECM.....	40
BILAGA 4. Programscript som tar fram prognoser för en specifik samplingsperiod	42

1. INLEDNING

Vi inleder med en diskussion om transaktionskostnader som kan uppstå vid applicering av prognosmodeller. Vi går igenom vår strategi för att inom uppsatsens ramar minimera dem. Därefter diskuterar vi antagandet om en starkt effektiv marknad och vilka prognosmöjligheter som kan uppstå vid ineffektivitet. Vi granskar resultatet av vår undersökning av tidigare gjorda arbeten inom prognostisering och lyfter fram syftet med vår uppsats. Kapitlet avslutas med en genomgång av arbetets disposition.

Transaktionskostnader är ett vanligt problem i samband med prognostisering. Om man kan påvisa prognoskraft så är det lätt att snabbt förklara bort den genom att lyfta fram transaktionskostnadsproblemet. Handelsstrategier som bygger på prognoser är ofta transaktionsintensiva. Detta är en viktig punkt eftersom transaktionskostnaden kan minimeras om man lyfter upp dem som ett centralt problem. Transaktionskostnader kan vara relaterade till handel men även till den arbetstid som går åt för beräknings- och analysarbete. Vi antar att vi kan få låga avgifter för handel och att transaktionskostnaderna istället har sin tyngdpunkt i arbetstid. Vi antar att de tidskrävande arbetsmomenten på marknaden uppstår i samband med inhämtning av information, prognosberäkning, analys och justering av portföljvalet. Vi antar också att de finansiella aktörerna vill ha möjlighet att justera sin position på kort varsel. T ex på intradag-data.

Vi ville redan från början minimera transaktionskostnaderna. För att öppna upp möjligheten så valde vi att genomföra vår prognosstudie med hjälp av ”programtrading-makro”. Alternativet hade varit att bygga matrisberäkningar i excelkalkylblad. Fördelarna med makro är inte uppenbara, men under arbetets gång så fick vi flera fördelar pga valet av prognosberäkningsstrategi. I uppstarten av vårt uppsatsarbete så var intentionen att vi skulle hämta data i realtid via internet. Anledningen var att vi med 100% sannolikhet ville säkerställa att våra prognoser inte använde sig av delar av tidsserierna som inte skulle ingå i prognosen (framtida värden). Vi kompromissade och valde istället att utveckla makro som beräknade prognoser baserat på historisk data i kalkylblad. Beräkningarna är fortfarande flexibla och minimerar vår arbetstid med att

göra beräkningar, utvärderingar och justera portföljer i samband med den simulering vi kommer att genomföra. Riskerna finns fortfarande att egenskaper hos makrona eller hos excel resulterar i att ”för ny” information används.

En del finansiella aktörer antar att marknaden är starkt effektiv. All information är känd av alla. Det gäller historisk data, aktuell data, diskonterad data och insiderinformation. I ett sådant sammanhang finns det inget utrymme för prognosmodeller. Trots detta så finns det svenska och internationella studier som påvisar prognoskraft. En uppenbar anledning är att marknaden är svagt effektiv och att någon eller några av kriterierna för stark marknad inte uppfylls. T ex antagandet om att alla aktörer får fram samma diskonterade värden för framtida observationer. Vi är intresserade av de teorier som antar att informations-spridning sker gradvis mellan olika segment av marknaden. Framför allt användningen av size-komponenten.

Om vi definierar information som nyheter och vi definierar parterna som privata eller institutionella placerare så kan vi anta att de tillgångar som handlas mest frekvent prisjusteras snabbast. Om en aktie handlas mindre frekvent så spelar det ingen roll om aktörerna har tillgång till all information, priset kommer inte att justeras förrän handel genomförs. Vi tänker oss stora företag med aktier med hög omsättning, *Big Cap*. Om det skulle vara så att små företag med lägre omsättning får sina aktierpriser justerade i en senare tidpunkt så öppnar sig möjligheten att marknaden inte är starkt effektiv och att det finns en prognostiseringsmöjlighet.

Vi undersöker möjligheten att prognostisera *Small Cap* avkastningar genom att använda oss av ett potentiellt samband över tiden mellan *Small Cap* och *Big Cap* index. En naturlig början är att arbeta med korsautokorrelationer mellan *Big Cap* och *Small Cap*. Korsautokorrelation därför att vi vill prognostisera ett aktieslag baserat på informationen om ett annat aktieslag i en tidigare tidpunkt.

De internationella studierna vi granskade har baserats på lead-lag effekt. Speciellt Lo & MacKinlay (1990) bearbetar lead-lag effekter. De fann lead-lag effekt, speciellt mellan stora och små bolag. Med hjälp av ett korsautokorrelationstest demonstrerade de även

att det är storbolagen som leder småbolagen men inte det motsatta. Lead-lag effekter har även dokumenterats av andra författare så som t ex Badrinath, Kale & Noe (1995) vilka går ett steg längre och undersöker komplexiteten i relationen. Andra forskare såsom Jegadeesh & Titman (1995) har använt diverse olika *contrarian* trading strategier, dvs de applicerar strategier som bygger på att marknaden har ett visst medelvärde och de köper de aktier som fallit och säljer de aktier som har stigit. De kom fram till att vinstmöjligheterna till stor del beror på överreaktion på marknaden och endast en mindre del på lead-lag effekter. Det intressanta för oss är att de hävdade att det finns lead-lag effekter.

Mark Arbeter (2006) argumenterar i sitt arbete, "Will Big Caps Back Down", att *Big Cap* är de som faller sist i en avmattande högkonjunktur. Om Arbeter har rätt så finns det korsautokorrelationsinformation mellan *Big Cap* och *Small Cap* som kan användas i en korsautokorrelationsmodell. Altay Erdinç (2004) undersöker korsautokorrelation mellan *Small Cap* och *Big Cap* på de tyska och turkiska marknaderna. Även Richardson Terry, Petersson David (1999) visar på samband i "The cross-autocorrelation for *size-based* portfolio returns is not an artefact of portfolio autocorrelation". Mot dessa står Speidell, Lawrence och Graves som påvisar att *Small Cap* ska ha en låg korrelation mot *Big Cap*. Att *Small Cap* och *Big Cap* har låg korrelation kan bero på Arbeters teori att de beter sig olika i olika faser av konjunkturcykeln. Det kan finnas en gemensam trend men samtidigt uppstå avvikelser pga konjunkturrörelser. Vi har bara nämnt en bråkdel av all den forskning som baseras på att det finns reaktionsskillnader och fördröjningar mellan stora och små bolag och att marknaden behandlar dem olika.

Vårt syfte med uppsatsen är att bedöma om det finns en gradvis informationsspridning mellan *Big Cap* och *Small Cap* på den svenska aktiemarknaden. Vi behöver tidsserier som kan fungera som substitut för indexen. Carnegie har ett småbolagsindex och stockholmsbörsen har ett storbolagsindex. Vår metod baserar sig på prognoser. Vår hypotes är att *Big Cap* leder *Small Cap* och att vi därmed kan prognostisera småbolagsindexet med hjälp av historisk information om bl a storbolagsindexet. Eftersom vi har två tidsserier så var det naturligt att använda bivariata modeller. Då vi använder historisk och nutida information hos en tidsserie för att prognostisera framtida

information om en annan tidsserie så använde vi oss av korsautokorrelationssamband. Om våra prognoser utvärderas att ha prognoskraft så kan vi alltså påvisa en gradvis informationsspridning mellan storbolag och småbolag.

Vår disposition av uppsatsen är uppdelad i *Data (2)*, *Prognosmetod (3)*, *Resultat (4)*, *Slutsats (5)*, *Bilagor (6)*, *Ordlista (7)* och *Referenser (8)*. I *Data (2)* så illustrerar vi de index som vi använt som substitut för värdet på svenska småbolag och svenska storbolag. Vi tar upp problematiken med likviditeten i småbolagsindex. Vi visar också hur vi använder oss av prognostiseringsfönster för att steg för steg ta fram prognoser. Vi tar också upp hur realtidsapplikationen påverkar vårt behov av data och dataströmmar. Vi undersöker visuellt om det verkar finnas en gemensam trend hos de två tidsserierna och vi avslutar med att ta upp eventuella tidsserier som exkluderats arbetet.

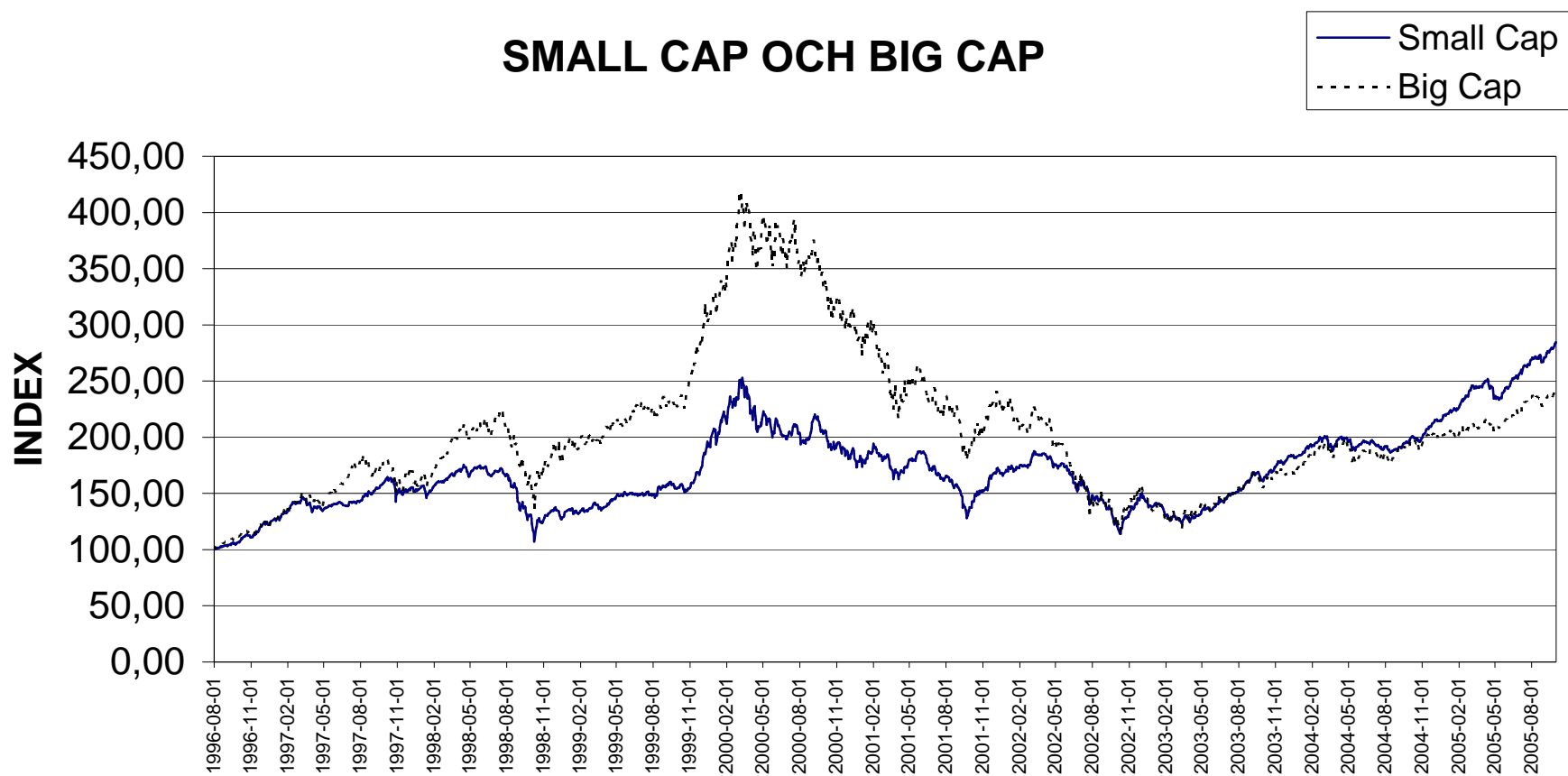
I *Prognosmetod (3)* så tar vi upp prognosteorin som för oss omfattar tidsserieegenskaper, estimeringsfönster, in-sample och out-of-sample prognoser, prognosmodellering, estimering av antalet laggade variabler, prognosutvärdering med hjälp av ME, MAE och RMSE, samt prognosutvärdering med hjälp av simulerade handelsstrategiresultat. I kapitlet går vi även igenom den *setup* som krävs för att genomföra prognosmetoden. Vi går igenom prognosmodellerna som ingår samt de referensmodeller som vi jämför mot. De bakomliggande beräkningar går igenom i detalj. Utvärderingsmetoderna går igenom och kapitlet avslutas med att gå igenom det bedömningskriteriet som vi valt att använda för att bedöma om prognoskraft föreligger.

Resultat (4) illustrerar de prognoserna i grafer tillsammans med det verkliga utfallet. Prognosutvärderingsresultaten redovisas och de bedöms enligt kriterierna definierade i *Prognosmetod (3)*.

Slutsats (5) sammanfattar *inledningen (1)*, *data (2)*, *prognosmetod (3)* och *Resultat (4)*. Valet av arbetsmodell kommenteras.

Bilagor (6) listar de bilagor som vi valt att bifoga uppsatsen. *Ordlista (7)* beskriver de ord och termer som vi ansåg behövde en extra förklaring. *Referenser (8)* listar de arbeten som vi har studerat i samband med uppsatsarbetet.

2. DATA & DATASTRÖMMAR



Figur 1. Small Cap och Big Cap i nivå

I de forskningsrapporter som vi har studerat så har forskningen utförts mestadels på amerikansk och engelsk data. Vi kommer att undersöka för lead-lag effekt mellan storbolag och småbolag på den svenska marknaden. Vi kommer att använda OMX30 index för att agera *substitut* för storbolagen. OMX30 är Stockholmsbörsen egna index för dess storbolag. Stockholmsbörsen har inget riktigt småbolagsindex, utan det referensindex som används av marknaden är sammansatt av fondbolaget Carnegie. Det heter CSX (se figur 1).

Ett problem med småbolag är att likviditeten kan vara begränsad och att index av den anledningen visar på förskjutningar i kursinformation pga att det inte sker kontinuerlig handel. Detta är dock inte fallet för Carninge CSX. Idag så sker det dagsvis flera affärer även för de mindre bolagen. Små bolag har generellt sämre likviditet än de stora. Dock sker det kontinuerlig handel i nästa alla bolag på Stockholmsbörsen och det finns endast ett fåtal riktigt småbolag som det förekommer en dag av icke-handel. Dessa fåtalet riktigt små bolag kvalificerar sig dock inte att ingå i O-listan (lista för Stockholmsbörsens mindre bolag då vi började skriva arbetet) utan de riktigt små bolagen återfinns på andra listor såsom NGM-listan och aktietorget. De likviditetsproblem som finns för de mindre bolagen idag är begränsat till att det är svårt att köpa och sälja större poster. Ett annat viktigt påpekande är att kurserna som vi använder är slutkurser.

Vi använder oss av ett rullande estimeringsfönster på t ex 500 dygn för att estimerar en prognostiseringsmodell och lägga en prognos för det nästkommande dygn (vi utvärderar även andra storlekar på prognosfönster). Sedan flyttar vi estimeringsfönstret ett dygn framåt i tiden och gör om proceduren. Vi använder oss av excelmakro för prognosberäkningen. Excelmakrot har tillgång till CSX och OMX30 i ett kalkylblad. Vi ville egentligen läsa börsdataströmmarna från internet för att kunna beräkna prognoser i realtid. En sådan lösning hade inneburit att vi inte haft den fasta datamängden på ca 2300 värden. Vi hade istället kontinuerligt fått in information. Detta är anledningen till att storleken på ca 2300 värden inte är viktig för oss. Alla beräkningar baserar sig på rullande estimeringsfönster med olika storlekar på samplingsintervallet.

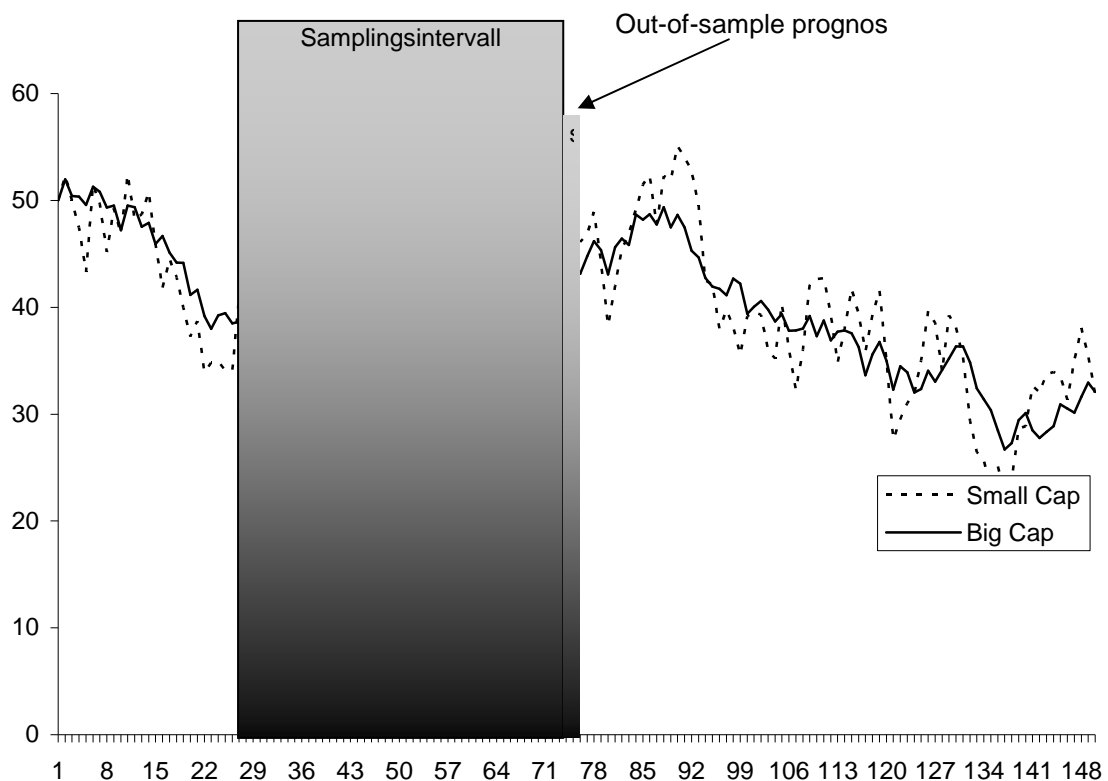
I nivådiagrammet kan vi se en tydlig trend. Vi ser också att de två tidsserierna är uppenbart individuellt icke-stationära. Att Small Cap skulle förlora värde snabbare än Big Cap i en börsnedgång kan vi inte se tecken på. Tvärtom verkar Small Cap gå bättre på sikt. Volatiliteten verkar vara större för *Big Cap*. En förklaring till att Small Cap går bättre på sikt kan vara att marknaden diskonterar en högre avkastningsfaktor på mindre bolag.

Storleken på indexen bestäms som summan av aktiekurserna multiplicerat med andel för respektive bolag i portföljen. Aktiekurserna sätts som de förväntade framtida vinsterna diskonterade med en specifik diskonteringsränta som varierar bolag mellan bransch. Informationsflödet påverkar värderingen relaterad till de framtida förväntningarna vilket är vad vi kommer att undersöka. T ex så kan informationen sprida sig med olika hastigheter till de små och de stora bolagen. Den andra faktor som styr aktiekursen förutom de förväntade framtida vinsterna är vilken diskonteringsränta som marknaden väljer att använda. En begränsning i vår data är att vi har exkluderat räntan.

3. PROGNOSMODELL

Vi inleder detta kapitel med ett prognosteoriansnitt. Vi förklarar olika prognosbegrepp som t ex estimeringsfönster (prognosfönster), in-sample-prognoser, out-of-sample prognoser, valet av antalet laggade förklarande variabler, univariat, bivariat, utvärderingskriterier och simulering. Därefter illustrerar vi vårt val av *setup* för prognosexperimentet. Prognosmodellerna går igenom grundligt och de bakomliggande beräkningarna visas steg för steg. Utvärderingskriterierna som indikerar prognosmodellernas *fit* beskrivs. Prognosutvärdering görs även med hjälp av simulering av handelsstrategier. Modellerna, hur de ska tolkas och hur de ska utvärderas beskrivs. Som avslutning på kapitlet så definierar vi de övergripande kriterierna som gäller för att vi ska kunna dra slutsatser på om *Big Cap* leder *Small Cap* på den svenska marknaden.

3.1 Prognosteori



Figur 2. Rullande samplingsintervall samt out-of-sample prognoser

Estimeringsfönster är den delmängd av de tillgängliga tidsserierna som används för att estimerera modellerna. Andra namn på samma sak är prognostiseringsfönster och samplingsintervall (se figur 2). Inom prognostisering så beräknar vi fram prognoser för en tidpunkt åt gången. Vid beräkningarna så används ett antal observationer. Man kan tänka sig att det är som ett fönster som successivt flyttas fram över en tidseriegraf. Ett steg för varje prognos. Antag att vi har 10 000 observationer. Vi vill bedöma hur bra prognosen för observation 5001 är. Vi bestämmer oss för att vi vill ha ett prognosfönster på 500 observationer. I vår estimering så använder vi oss av en sampleperiod på 500 värden som börjar vid observation 4501 och slutar på observation 5000. Vi beräknar prognosen för nästa tidpunkt och sedan kan vi jämföra med det verkliga utfallet för observation 5001. Skillnaden mellan det verkliga utfallet och prognosen kallas prognosfel. I vårt fall så arbetar vi med out-of-sample prognoser. Det betyder att prognosen vi beräknar fram inte är inkluderad i sampleperioden. Vid in-sample-

prognoser så är det verkliga värdet för observationen i prognospunkten inkluderat i sampleperioden. Vi bedömde det som en onödig risk att använda oss av det verkliga värdet när vi beräknade prognoserna och använde uteslutande out-of-sample prognoser.

Vid val av lag så används normalt *Akaike informations Criteria* eller *Swarts*. Vi valde dock godtyckligt en dags fördröjning. Vi antar att om det sker någon form av informationsförskjutning så kommer vi även att se det på dags fördröjningen. Vi hade mycket väl kunnat välja en modell med två dagars lag, eller en modell där flera dagar har påverkan på resultatet, eller i bästa fall testat en rad av modeller och sedan valt den bästa. Vi ville dock inte att redovisa ett bra resultat givet vårt val av data. Vi vill skapa en godtycklig modell och se om den har prognostiseringsstyrka. Ett allvarligt problem med att optimera den data man har är att man hittar den modell som ger bästa möjliga resultat och drar slutsatser därefter. Men slutsatserna behöver inte vara korrekta i ett annat sammanhang eftersom modellen var anpassad för just den data man hade. Därför är vi nöjda med att välja ej optimerade modeller med en godtycklig lag på 1 dag.

En univariat tidsseriemodell kan användas för att modellera och göra prognoser på en finansiell variabel baserat på dess historiska värde och det historiska värdet på dess felterm. En tidsseriemodell är en a-teoretisk modell, dvs inte baserad på en teoretisk modell för en variabel utan den försöker istället att fånga empiriskt relevanta egenskaper hos observerade data. En bivariat modell används då man har två tidsserier där egenskaper från den ena tidsserien kan påverka den andra. Vi kommer att använda bivariata modeller då vi har två tidsserier, *Small Cap* och *Big Cap*.

Utvärderingskriterierna för en prognosmodell kan basera sig på t ex prognosfel. Prognosfel är skillnaden mellan prognostiserat värde och verkligt utfall. Ett utvärderingskriterium är Mean Absolute Error (MAE). Man beräknar det som medelvärdet på absolut-progوسفelen i en tidsserie av prognoser. Just detta mått ska vara så nära noll som möjligt. Om värdet är noll så har man perfekt *fit* enligt kriteriet. Felet mellan prognos och verkligt utfall är med andra ord då i genomsnitt noll.

Prognosutvärdering kan även göras med simulering av handelsstrategier. Detta gör man om man vill utvärdera den ekonomiska signifikansen hos en prognosmodell. T ex så kan man sätta upp en handelsstrategi som agerar på ”sign”. Dvs vid positiv prognos så köper man och vid negativ prognos (eller noll) så säljer man.

Gemensamt för både prognosutvärderingsmetoderna ovan är att det krävs referensprognosmodeller. För MAE så är *fit* är ett relativt begrepp. Om *fit* är 0,25 så vet man inte om det är bra eller dåligt. Om man har en referensprognosmodell som ger ett MAE på 0,70 så vet man i alla fall att man har bättre *fit* på prognosmodellen än på referensprognosmodellen. För simulering med handelsstrategier så kan man få en vinst på 6 000 kr under en vald tidsperiod. Men hur vet man om detta är bra eller dåligt? Om man har en referensmodell som ger en vinst på 4 500 kr på samma tidsperiod så går det att dra slutsatser.

3.2 Setup av prognosmetod

Vi ger här en kortfattad beskrivning hur vi satta upp prognosexperimentet. Beskrivning är ämnad att ge en överblick på de ”komponenter” som behövs för genomförandet. Komponenterna beskrivs individuellt i de nästkommande avsnitten.

Vi valde att använda oss av out-of-sample prognoser och ett rullande prognostiseringsfönster. Till en början så valde vi en fast fönsterstorlek på 500 observationer. En fullängdsestimering med fast startpunkt var inte aktuellt. Det hade inneburit att tidiga prognoser hade fått värden baserade på färre observationer än de nyare prognoserna. Dessutom så hade vi frångått iden om att utveckla makron vilka kan fungera i realtidssammanhang där det inte finns en start och ett stopp för tidsserierna. Ett för kort sample hade försämrat kvaliteten i den statistiska regressionen. Ett för långt sample hade försämrat antalet prognoser. Om vi t ex hade använt 2000 sample så hade vi bara fått 97 prognoser. Ett helt naturligt val i vårt fall var att välja 500 sample. Då får vi över 1500 prognoser. 500 sample som motsvarar nästan 1,5 år är också en intressant samplestorlek eftersom den följer konjunkturen. Den fångar antingen en uppgång, en nedgång, en del eller en topp. När man gör en prognos för nästa tidpunkt i en högkonjunktur så verkar det logiskt att bara använda ett prognossample för den typen av

konjunktur som man befinner sig i. Vi fann att vår argumentation var värd att ifrågasättas. Vi adderade prognosintervall i hela skalan från 100 till 2000 med steg om 100 observationer. Vi fick upprepa hela vårt prognosexperiment 20 gånger. Här fick vi nytta av vårt makro. Efter ett par justeringar så fick vi fram prognoser för samtliga valda längder på prognosfönstret. Om man gör sina prognoser med hjälp av ett realtidssystem med automatiska köp- och sälj-script så finns möjlighet att låta makrot löpande beräkna den optimala längden på estimeringsintervall baserat på historiska prognosresultat. Prognoser handlar om att använda alla medel tillättna för att finna bästa estimat för nästa tidpunkts aktiekurs.

Vi valde två bivariata prognosmodeller. Den första är en VAR modell som baserar sig på korsautokorrelation mellan *Big Cap* och *Small Cap*. Den andra är en VECM-modell som baserar sig på samma samband fast med en inkluderad felkorrektionsterm. Vi använde dygnsdata. Prognosutvärderingarna genomfördes med dels de tre kriterierna ME, MAE och RMSE och dels med simulering av handelsstrategier. För varje storlek på prognosfönster så genomfördes samtliga typer av utvärderingar. Vi valde referensmodeller som prognosmodellerna måste slå för att visa på prognoskraft. Till ME-, MAE- och RMSE-utvärderingen valdes en Random Walk prognosmodell. Till simuleringen valdes en passiv buy-and-hold strategi med 100% placering i *Small Cap*. Det övergripande utvärderingskriteriet var att prognosutvärderingarna för de bivariata prognosmodellerna skulle vara bättre än referensmodellerna.

3.3 Val av prognosmodeller

Givet att både storbolag och små bolag är informationsstyrda så ska de röra sig åt samma riktning vid samma typ av information. Det vi vill undersöka är om det sker en fördröjning i informations spridning till småbolag. Ett sätt att undersöka för fördröjning i informations spridning är att testa ifall småbolagen följer storbolag. Då antar vi att småbolagen följer storbolagen av den anledningen att först reagerar storbolagen på ny information, därefter reagerar småbolagen. Vi har valt bivarata modeller av den anledningen att vi har två tidserier, där prognostiseringen ska vara en funktion av bägge tidsseriernas föregående värde. Den första modellen är en bivariat VAR-modell som

kan fånga upp korsautokorrelation mellan tidsserierna. Den andra bivariata prognosmodellen baserade vi på iden med den långsiktiga trenden och en felkorrigering på kort sikt. Vi använder en enkel bruskorrigering (ECM) likt den Engel Granger fick Nobelpris för. Den modell vi fick fram är en VECM (Vector Error Correction modell).

3.4 Prognosmodell 1 - VAR

En VAR-modell (Vector Auto Regressive) beskriver en uppsättning endogena variabler över en sample-period som en funktion av deras tidigare utveckling (Enders 2003). Den kan skrivas i differensform och den kan vara bivariat.

$$\Delta y_t = a_{10} + a_{11}\Delta y_{t-1} + e_{yt} \text{ (AR modell i differensform, univariat)}$$

$$\Delta y_t = a_{10} + a_{11}\Delta y_{t-1} + a_{12}\Delta x_{t-1} + e_{yt} \text{ (Bivariat VAR – Prognosmodell 1 - VAR)}$$

$$\Delta x_t = a_{20} + a_{21}\Delta y_{t-1} + a_{22}\Delta x_{t-1} + e_{xt}$$

Small Cap representeras av y_t och *Big Cap* av x_t . Vi använder nivåvärdena på indexserierna för att få samma struktur på modellen som hos *prognosmodell 2 (VECM)*. Differenser använder vi av samma anledning. Den bivariata modellen har två ekvationer och vi använder den första av dem för våra VAR prognoser. Ett exempel på en bivariat modell är den som utvecklades av Näsman i en tidigare uppsats, Näsman (2003). I denna uppsats användes en vektor av branschindex (endogena variabler) för att prognostisera nästa tidpunkts index.

3.5 Långsiktiga samband

Ett antagande som vi har gjort är att eftersom vi inte har en branschdiversifiering utan istället en småbolags- och storbolagsdiversifiering så kan man anta att indexen bör representera den ekonomiska aktiviteten som helhet och inte en specifik bransch. Givet detta antagande så bör småbolags- och storbolagsindex ha ett långsiktigt samband, vilket är logiskt då positiv information för ekonomin bör påverka både de små och de stora bolagen åt samma riktning.

Teoretiskt så har vi gjort antagandet att de båda tidserierna har ett långsiktigt samband med varandra. Vi har gjort antagandet att serierna är kointegrerade. Dvs de har en stationär relation mot varandra eller har ett långsiktigt samband med men är ej fullt ut stationärt.

Dvs antagandet är att de bägge tidserierna ska (har) starka kointegrations tendenser. Två tidserier är kointegrerade då det existerar ett värde β där $Y_t - \beta X_t = Z_t$ och Z_t är $I(0)$ dvs stationär mot varandra som innebär att de är integrerade av graden noll eller med andra ord att de bägge tidsserierna följer en gemensam trend.

I de fallen då båda tidsserierna är integrerade av första graden och de bägge tidsseriernaserierna är kointegrerade med varandra får man en kointegrationsvektor (Z_t) av typen $(1, -\beta)$. Det är viktigt att Z_t inte är $I(1)$. I det fallet som Z_t är $I(1)$ adderar man visa parametrar för att sluta producera Z_t . Syftet är att tillsätta rätt parametrar så att Z_t inte blir integrerad av första graden vilket skapar begränsningar för den långsiktiga jämvikten mellan de bägge tidserierna. Förenklat kan man uttrycka att då man har ett Z_t som är $I(0)$ så är tidserierna kointegrerade.

I det fallet som kointegrationskoefficienten konvergerar i en snabbare takt mot dess jämvikt än mot andra konventionella asymptot kallar man kointegrationskoefficienten för superkonsistent. En superkonsistent koefficient underlättar vid prognostisering, då större avvikelser från jämvikten förväntas plana ut fortare.

I detta arbete behandlar vi två tidsserier, småbolagsindex och storbolagsindex. Dessa två tidsserier är bägge två oberoende *RW tidsserier*, dock så agerar de på samma information och de kan därför ses som att de har ett långsiktigt samband, detta fenomen kallas *Spurious regression*, dvs man har två oberoende tidsserier av typen

$$Y_t = Y_{t-1} + \varepsilon_{1t}$$

$$Y_t = Y_{t-1} + \varepsilon_{2t}$$

Där ε_{1t} och ε_{2t} har ett förhållande med varandra.

I vårt fall har ε_{1t} och ε_{2t} eftersom bägge reagerar på samma information men av olika grad beroende på typen av informationen. Logiskt är att småbolagen reagerar starkare

på företagsspecifik information som gäller storbolag än vad storbolag reagerar på företagsspecifik information för småbolag. En förklaring kan vara att småbolag kan vara underleverantörer till storbolagen och därav är de direkt beroende av den information som rör storbolagens utveckling medan storbolag reagerar mindre på information från småbolag.

3.6 Prognosmodell 2 - VECM

Vi antar att våra tidsserier delade en gemensam långsiktig trend. I det korta perspektivet så skiljer sig *Small Cap* mycket från *Big Cap*. Om vårt antagande om gradvis informationsspridning stämmer så kan det finnas andra faktorer som påverkar ett eventuellt samband på kort sikt. Vi blev intresserade av Engel och Grangers applikation där de föreslår förklaringar på långsiktiga gemensamma trender och skillnader i det korta perspektivet. Vi utgick från Engel och Grangers tre stegsmodell och utvecklade en bivariat prognostiseringsmodell som inkluderar en felkorrigeringsterm. Engel och Grangers trestegsmodell (Enders version) finns bifogad i bilaga 1.

$$y_t = a + bx_t + e_t \text{ (Engel-Granger Steg 2)}$$

$$\Delta y_t = a_{10} + a_y \hat{e}_{t-1} + a_{11} \Delta y_{t-1} + a_{12} \Delta x_{t-1} + e_{yt} \text{ (Engel-Granger Steg 3)}$$

Engel-Granger Steg 3 använde som *Prognosmodell 2 (VECM)*. För att kunna göra detta så krävdes det att vi tog fram de estimerade residualerna med hjälp av *Engel-Granger Steg 2* ovan. I vår applicering så är y_t *Small Cap* och x_t är *Big Cap*. Estimeringen av residualerna ger \hat{e}_{t-1} vektorn som behövs för att beräkna prognoserna i *Prognosmodell 2 (VECM)* nedan.

$$E(\Delta y_t) = a_{10} + a_y \hat{e}_{t-1} + a_{11} \Delta y_{t-1} + a_{12} \Delta x_{t-1} \text{ (Prognosmodell 2 – VECM)}$$

Vi är intresserade av om felkorrektionen ger en förbättring av prognoskraften. Lyckas VECM bättre än VAR modellen så har vi även en indikation på att en felkorrigeringskomponent ger ett positivt prognostiseringstillskott.

Engel och Granger introducerade begreppet kointegration 1987. De definierar i sin analys ett *long-run equilibrium error*. En egenskap hos kointegrerade variabler är att deras tidsserier träffas av avvikelse från *long-run equilibrium*. Om systemet ska återgå till jämvikt så måste åtminstone några variabel reagera för att återställa balansen. En modell baserad på denna dynamik kallas felkorrigeringsmodell och arbetar med två horisonter, det korta och det långa perspektivet. På kort sikt påverkas modellen av avvikelser från jämvikt. På lång sikt så har vi jämvikt.

$$\Delta y_t = a_{10} + a_y(y_{t-1} - bx_{t-1}) + a_{11}\Delta y_{t-1} + a_{12}\Delta x_{t-1} + e_{yt}$$

$$\Delta x_t = a_{20} + a_x(y_{t-1} - bx_{t-1}) + a_{21}\Delta y_{t-1} + a_{22}\Delta x_{t-1} + e_{xt}$$

Där a är riktningskoefficienter för dess förknippade par. Feltermerna och differenstermerna är stationära och den linjära kombinationen av $(y_{t-1} - bx_{t-1})$ måste vara stationär. Vi får en bivariat VAR modell med förstaddifferenser som är förstärkt med en *Error Correction* term $a_y(y_{t-1} - bx_{t-1})$ och motsvarande term för den andra ekvationen. a_y och a_x är *speed of adjustment term*. Ju större de är desto större är svaret på förra periodens avvikelse från *long-term equilibrium*. Små termer betyder att Δy_t och Δx_t inte reagerar på tidigare periodens *equilibrium error*. Om delta Δy_{t-1} ska vara opåverkad av Δx_{t-1} så måste a_y och a_{12} vara noll. En av de två *speed-of-adjustment* termerna måste vara skild från noll. Annars får vi inte *long run equilibrium* sambandet och vi får inte en felkorrigeringsmodell eller kointegration.

Om a_y och a_x är noll så har vi en traditionell VAR med förstaddifferenser. Ingen felkorrigering kan representeras. Δy_t reagerar då inte på tidigare avvikelse ifrån *long-run equilibrium*. Om en av a_y och a_x skiljer sig från noll så reagerar Δy_t på tidigare periods avvikelse från jämvikt. Att använda en VAR modell hade resulterat i felspecificering eftersom y_t har en felkorrigeringsrepresentation.

3.7 Referensprognosmodell – Random Walk

Vid *prognosutvärdering 1 (Prognosutvärderingskriterier)* så får vi fram prognosutvärderingskriterier i form av ME, MAE och RMSE. Dessa kriterier anger prognosernas *fit* mot det verkliga utfallet. En bra *fit* indikerar prognoskraft. Ett enkelt antagande är att kriterierna för prognosmodellerna ska ha bättre *fit* än en *Random Walk prognosmodell*. Vi utvecklade en sådan referensmodell och denna använder vi för att jämför våra resultat mot. Referensmodellen finns även med i *utvärdering 2 (Simulering av handelsstrategier)*. Där exkluderas den dock från analysen. En annan referensmodell används.

$$E(\Delta y_t) = \Delta y_{t-1} \text{ (Random Walk referensprognosmodell)}$$

Den första termen (V.L) är det förväntade värdet av differensen i index för *Small Cap*. Dvs $E(y_t - y_{t-1})$ där y_t är *Small Cap*. Den andra termen (H.L) är differensen i index för *Small Cap* för den föregående tidpunkten. Dvs det senast kända värdet för differensen. Detta är alltså en mycket enkel prognosmodell som antar att den bästa förutsägelsen av nästa tidpunkts differens i *Small Cap* är det tidigare kända värdet på samma differens. Anledningen till att vi använder differenser är att vi vill ha samma struktur på modellen som den som utvecklats för *prognosmodell 1 (VAR)* och *prognosmodell 2 (VECM)*

3.8 Val av transformation

Big Cap och *Small Cap* är indextidsserier. Tidsserierna normaliseras så att de i första tidpunkten har värdet 100. Vi valde mellan att arbeta med avkastningar eller differenser. Engle och Grangers trestegsmetod som vi använde som bas för att utveckla vår VECM prognosmodell använder sig nivåtidsserier i steg 1. De bedömer i steg 1 om tidsserierna är I(1). I steg 2 så använder de samma tidsserier (i nivå) för att ta fram residualer. Residualerna används i steg 3 för att ställa upp en modell som lätt kan skrivas om till en prognosmodell. Se ekvation 2 i bilaga 1. Vi gjorde därför valet att arbeta med differenser på våra index. Då vi tar fram residualerna så använder vi oss av

index för *Small Cap* och *Big Cap* i nivå. Då vi går vidare och använder residualerna i *Prognosekvation 2 (VECM)* så gör använder vi differenserna för *Small Cap* och *Big Cap* index. Vi arbetar inte med procentuella avkastningar någonstans i arbetet. Vi valde därför att inte arbeta med logaritmerade värden utan vi utgick istället ifrån indexserierna. Om inte VECM hade varit vår huvudprognosmodell så hade vi valt annorlunda.

3.9 Utvärdering 1 - Prognosutvärderingskriterier

Mått	Beskrivning	Syfte	Beräkning	Utfall
ME	Mean Error (Mått på överskattning/underskattning)	Visar om modellen i genomsnitt över- eller underskattar observerade avkastningar	$\frac{1}{N} \sum e_t$	Neg. = underskattar Pos. = överskattar
MAE	Mean Absolute Error (Mått på prognosfelets storlek)	Beräkning av den absoluta skillnaden mellan prognos och observation. En bra prognos med bra träff ger ett värde nära 0.	$\frac{1}{N} \sum e_t $	Nära 0 = bra fit
RMSE	Root Mean Square Error (Mått på prognosfelets storlek)	RMSE skiljer sig från MAE då större vikt ges åt de avvikelser som avviker mer än medelavvikelsen.	$\frac{1}{N} \sum e_t^2$	Nära 0 = bra fit

Tabell 1. Utvärderingskriterium

De första 500 observationerna används för att beräkna det första prognosvärdet. Av ca 2300 observationer så får vi ut ca 1800 prognoser. Prognosvärdena utvärderas med hjälp av prognosutvärderingskriterierna ovan. Felet mellan utfall och prognos definieras som e_t . Definitionen av prognosutvärderingskriterierna finner ni i tabellen ovan. Dessa kommer att användas vid utvärdering av vår huvudprognosmodell, VECM, samt utvärdering av våra två referensmodeller, VAR och RW. Vi upprepar metoden för samtliga storlekar på prognostiseringsfönster.

3.10 Utvärdering 2 – Simulering av handelsstrategier

Strategi	Beskrivning
S0	Passiv placering av investeringsbeloppen i Småbolagsfond.
S1	Investering i Småbolagsfond då VECM-prognos visar på en positiv differens. Annars placering i riskfri tillgång.
S2	Investering i Småbolagsfond då VAR-prognos visar på en positiv differens. Annars placering i riskfri tillgång.
S3	Investering i Småbolagsfond då RW-prognos visar på en positiv differens. Annars placering i riskfri tillgång.

Tabell 2. Handelsstrategier

Utöver prognosutvärderingskriterier så använder vi oss av en simuleringsmetod för att utvärdera prognosmodellernas ekonomiska signifikans. Vi skapar fyra stycken handelsstrategier (tabell 2). Det är det monetära utfallet av en investering som är intressant. Alla modellerna utgår från ett portföljvärde på en godtycklig summa. Vi valde 10 000 kr. Vi antog att transaktionskostnaden noll. Alla strategierna är av samma slag. Vi antar att vi kan placera i småbolagsindex och att en riskfri tillgång finns som vi kan handla med. Vi läser av prognosen av differensen i småbolagsindex $E(\Delta y_t)$. Om den är positiv så investerar vi 100% av kapitalet i småbolagsindex. Om den är negativ så investerar vi 100% av kapitalet i den riskfria tillgången (räntan antas vara 0%). Den första handelsstrategin är dock passiv. Dvs vi investerar hela beloppet i småbolagsfonden och låter dem ligga kvar i den tillgången. De övriga strategierna prognostiserar alla *Small Cap* och samma teknik används. Dvs för VECM, VAR respektive RW.

3.11 Övergripande bedömningskriterier

Prognoserna utvärderas mot referensmodeller. De bivariata prognosmodeller ska ge bättre *fit* och bättre avkastning i simuleringen med handelsstrategier. Om så är fallet så kan vi påvisa prognoskraft vid prognostisering av *Small Cap* med hjälp av *Big Cap*. Vi kan därmed också dra slutsatsen att det finns tecken på en gradvis informationsspridning mellan *Big Cap* och *Small Cap*.

4. RESULTAT

4.1 Inledning

I graferna i avsnitt 4.2 så sammanställer vi prognosernas $E(\Delta y_t)$ tillsammans med det verkligt utfallet Δy_t . De bivariata prognosmodellerna genererar vars en graf. Graferna är bara exempel. De illustrerar bara en sektion av tidsserien för en av storlekarna på prognosfönster (det finns 20 olika). De tre graferna tolkas för att få en uppfattning om prognosförmågan hos VECM-prognosmodellen i jämförelse med VAR- och RW-prognosmodellerna.

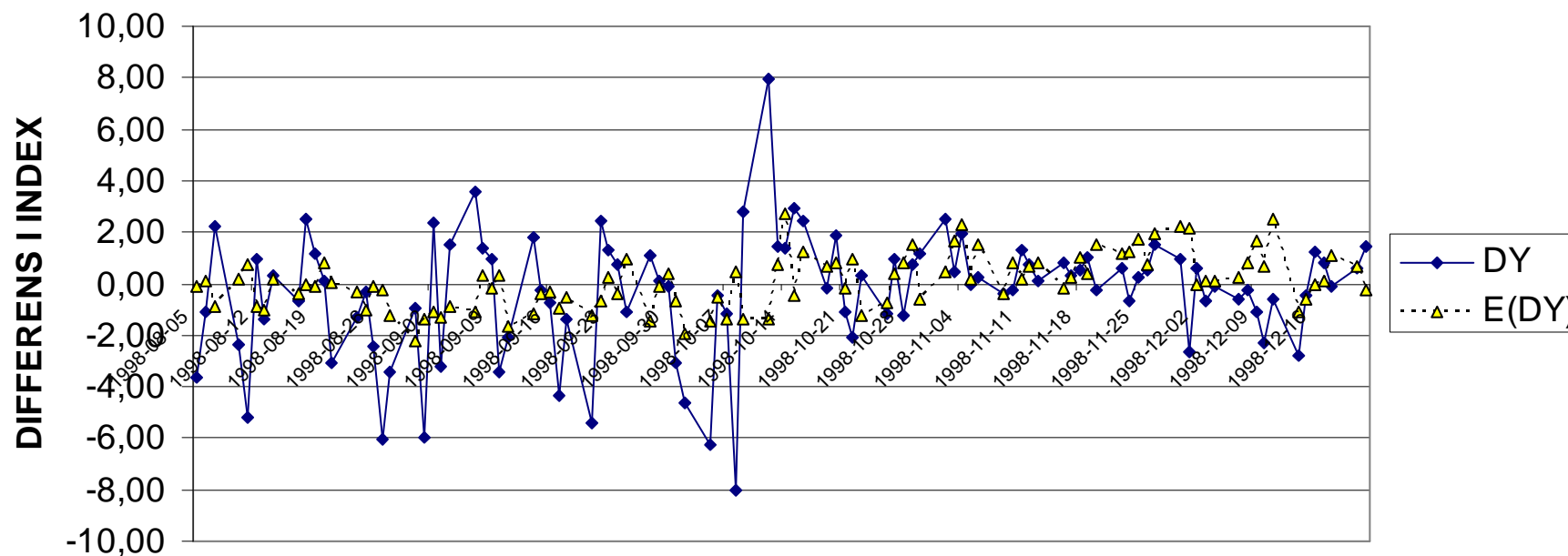
Därefter använder vi oss av prognosutvärderingskriterierna i tabellform. Även här är målsättningen att avgöra om det finns prognostiseringsförmåga hos de bivariata prognosmodellerna i jämförelse med referensprognosmodellen.

Den ekonomiska signifikansen analyseras genom en simulering. I respektive strategi så placerar vi 10 000 kr 1998-08-04 (tidpunkten innan den första prognosen vid prognosfönster på 500 observationer). Vi använder sedan handelsstrategierna listade i tabell 2. Strategi 0 (S0) baserar sig inte på en prognosmodell. Det är en referensstrategi där man låter pengarna stanna i småbolagsfonden genom hela simuleringsperioden. Tidpunkten för den första prognosen kom att skilja sig från den angivna datumen ovan i samband med att vi utökade metoden till att utvärdera andra estimeringsfönster.

Vi avslutat kapitlet med att applicera bedömningskriterierna som definierades i slutet av kapitel 3.

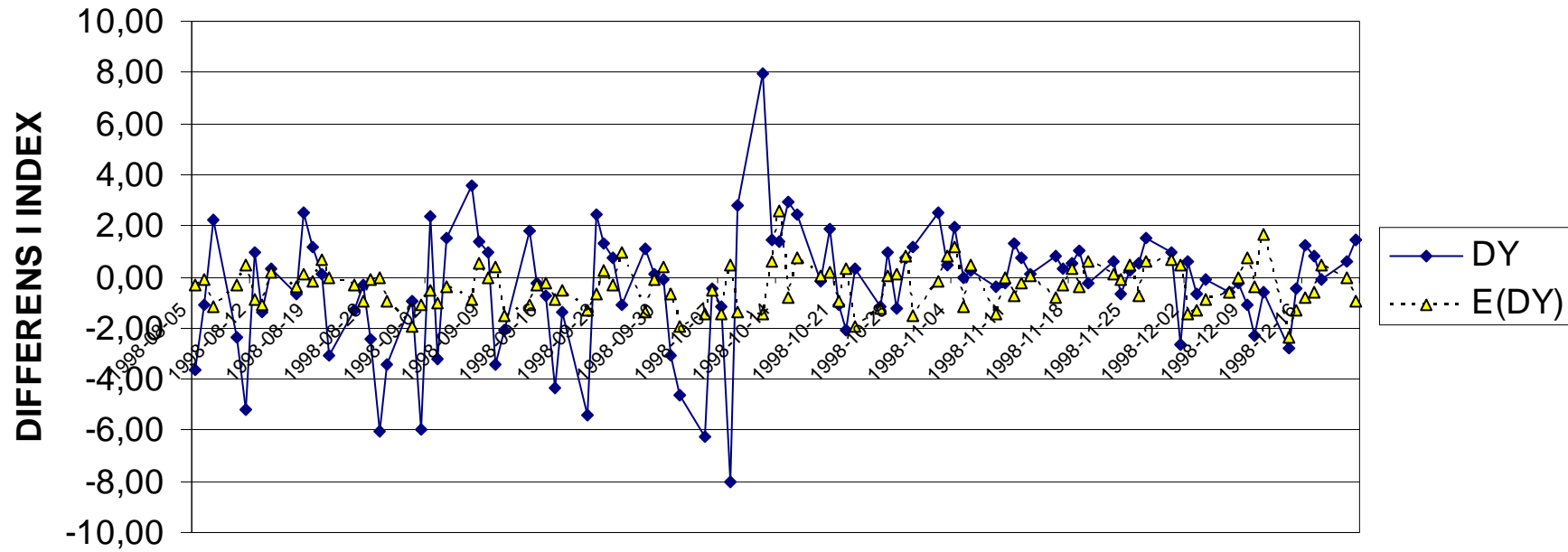
4.2 Grafer över prognoser i jämförelse med verkligt utfall

VECM-PROGNOSER (100 ST)



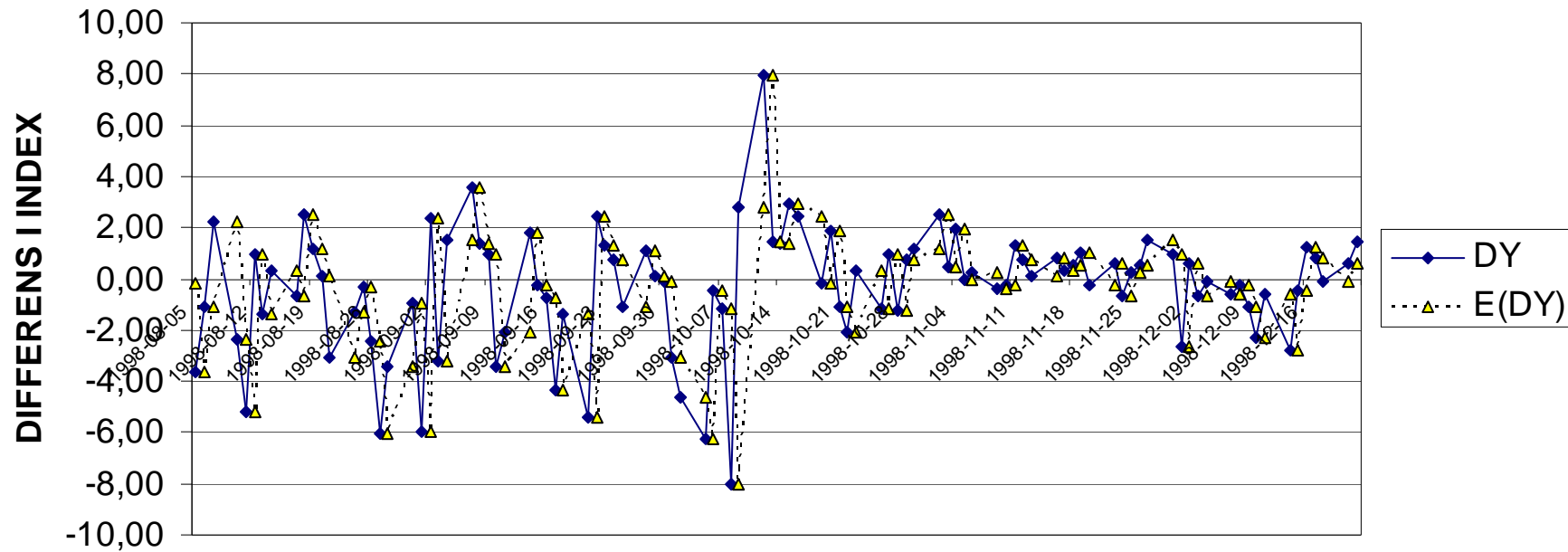
Figur 3. VECM-Prognoser, $E(DY)$, samt verkligt utfall, DY , för tidsperioden 1998-08-05 till 1998-12-16 och ett prognosfönster på 500 sample

VAR-PROGNOSER (100 ST)



Figur 4 VAR-Prognoser, $E(DY)$, samt verkligt utfall, DY , för tidsperioden 1998-08-05 till 1998-12-16 och ett prognosfönster på 500 sample

RW-PROGNOSER (100 ST)



Figur 5. RW-Prognoser, $E(DY)$, samt verkligt utfall, DY , för tidsperioden 1998-08-05 till 1998-12-16 och ett prognosfönster på 500 sample

VECM-prognoserna är mycket lika VAR-prognoserna. Prognosvärdena ligger genomgående lägre för VECM. RW-prognoserna ser ut som förväntat. Prognosen $E(\Delta y_t)$ får här värdet av Δy_t i föregående tidpunkt. I delar av grafen så är prognoserna bra, men det är svårt att avgöra prognosförmågan bara genom att granska figurerna.

4.3 Utvärdering 1 - Prognosutvärderingskriterier

Window	VECM			VAR			RW		
	ME	MAE	RMSE	ME	MAE	RMSE	ME	MAE	RMSE
100	-0,02	1,11	2,97	0,00	1,09	2,91	0,00	1,38	4,66
200	0,00	1,15	3,07	0,00	1,13	3,02	0,00	1,44	4,88
300	0,00	1,21	3,22	0,01	1,19	3,17	0,00	1,52	5,12
400	0,02	1,26	3,34	0,01	1,25	3,32	0,00	1,60	5,39
500	0,02	1,33	3,51	0,01	1,32	3,48	0,00	1,68	5,69
600	0,07	1,30	3,31	0,04	1,29	3,30	0,00	1,65	5,52
700	0,08	1,33	3,45	0,04	1,33	3,44	0,00	1,70	5,76
800	0,09	1,37	3,59	0,05	1,36	3,59	0,00	1,74	6,02
900	0,06	1,31	3,20	0,00	1,30	3,19	0,00	1,68	5,60
1000	0,10	1,19	2,52	0,04	1,18	2,51	0,00	1,52	4,31
1100	0,14	1,12	2,26	0,06	1,11	2,25	0,00	1,43	3,81
1200	0,15	1,06	1,98	0,07	1,05	1,96	0,00	1,35	3,35
1300	0,20	1,03	1,83	0,10	1,02	1,81	0,00	1,32	3,12
1400	0,18	1,02	1,79	0,08	1,00	1,77	0,00	1,28	2,97
1500	0,22	0,97	1,57	0,12	0,95	1,53	0,00	1,22	2,55
1600	0,25	0,92	1,39	0,15	0,89	1,34	0,00	1,15	2,29
1700	0,30	0,93	1,42	0,18	0,90	1,35	0,00	1,16	2,36
1800	0,29	0,95	1,48	0,17	0,92	1,42	0,00	1,20	2,53
1900	0,23	0,95	1,55	0,15	0,94	1,53	0,01	1,25	2,75
2000	0,28	0,93	1,46	0,23	0,92	1,43	0,00	1,17	2,28

Tabell 3. Prognosutvärdering i tabellform för VECM-, VAR-, och RW-prognoser för samtliga prognosfönsterstorlekar

I tabellen ovan så har vi prognosfönstrets storlek i kolumnen längst till vänster. Felet mellan prognos och verkligt utfall, e , samt antalet prognoser ger prognoskaraktärstiken för var och en av modellerna. Se beräkningsuppställningar i tabell 1 samt beskrivningarna nedan.

Mean Error (ME)

RW-prognos skattar varken över eller under det verkliga värdet. Värdet verkar ligga i genomsnitt på noll. Huruvida denna är stor eller inte är svår att säga. Både VAR- och VECM-prognosmodellerna överskattar det verkliga utfallet något.

Mean Absolute Error (MAE)

VAR-prognosmodellen ger det MAE-värde som ligger närmast 0, dvs den har bäst *fit*.
VECM ger ett liknande resultat men är sämre. RW-modellen visar upp sämst *fit*.

Root Mean Square Error (RMSE)

För VAR-prognos så får vi ungefär samma *fit* som för VECM-prognos. VAR-prognoserna har dock bättre *fit*. RW-prognos ger ett mycket sämre värde. RMSE ger större vikt till uteliggande prognoser. Om vi granskar grafen för RW-prognoser mot verkligt utfall så ser vi att amplituden för dessa prognoser ligger i nivå med verkligt utfall. När ett fel inträffar så blir det stort. Att RW-modellen får en dålig *fit* kan på så sätt förklaras. RW-prognosmodellen skiljer sig från VECM- och VAR-modellerna då den inte beräknas med regression.

Prognosutvärdering

VAR-modellen ser ut att kunna prognostisera bättre än VECM-modellen. RW-modellen som är referensmodell ger ett sämre *fit*.

4.4 Utvärdering 2 – Simulering av handelsstrategier

Win	VECM				VAR				RW				Buy-Hold
	ME	MAE	RMSE	Sim	ME	MAE	RMSE	Sim	ME	MAE	RMSE	Sim	Sim
100	-0,02	1,11	2,97	28024	0,00	1,09	2,91	29720	0,00	1,38	4,66	92753	23124
200	0,00	1,15	3,07	26864	0,00	1,13	3,02	26880	0,00	1,44	4,88	88307	20474
300	0,00	1,21	3,22	15149	0,01	1,19	3,17	19962	0,00	1,52	5,12	78723	17398
400	0,02	1,26	3,34	19252	0,01	1,25	3,32	22098	0,00	1,60	5,39	76145	17484
500	0,02	1,33	3,51	26890	0,01	1,32	3,48	20667	0,00	1,68	5,69	69324	17213
600	0,07	1,30	3,31	23484	0,04	1,29	3,30	24150	0,00	1,65	5,52	59849	22238
700	0,08	1,33	3,45	21505	0,04	1,33	3,44	20392	0,00	1,70	5,76	52549	19309
800	0,09	1,37	3,59	18504	0,05	1,36	3,59	19593	0,00	1,74	6,02	53301	18210
900	0,06	1,31	3,20	16231	0,00	1,30	3,19	15685	0,00	1,68	5,60	32717	12198
1000	0,10	1,19	2,52	17863	0,04	1,18	2,51	16798	0,00	1,52	4,31	33402	13628
1100	0,14	1,12	2,26	20980	0,06	1,11	2,25	17056	0,00	1,43	3,81	33883	15048
1200	0,15	1,06	1,98	19533	0,07	1,05	1,96	18119	0,00	1,35	3,35	31087	15937
1300	0,20	1,03	1,83	19134	0,10	1,02	1,81	18151	0,00	1,32	3,12	28757	20740
1400	0,18	1,02	1,79	16994	0,08	1,00	1,77	15955	0,00	1,28	2,97	26035	16250
1500	0,22	0,97	1,57	16377	0,12	0,95	1,53	18672	0,00	1,22	2,55	26976	19378
1600	0,25	0,92	1,39	14725	0,15	0,89	1,34	18072	0,00	1,15	2,29	22030	19878
1700	0,30	0,93	1,42	14333	0,18	0,90	1,35	16893	0,00	1,16	2,36	19801	20899
1800	0,29	0,95	1,48	13273	0,17	0,92	1,42	15213	0,00	1,20	2,53	16123	17556
1900	0,23	0,95	1,55	13256	0,15	0,94	1,53	13306	0,01	1,25	2,75	13807	14322
2000	0,28	0,93	1,46	13072	0,23	0,92	1,43	13806	0,00	1,17	2,28	13543	15043

Tabell 4. Simuleringsresultat

Vi analyserar utfallet av de fyra handelsstrategierna. Vi startar med fyra portföljer med 10 000 kr i vardera. Tidpunkten är 1998-08-04. Den är 500 dagar efter den första dagen i tidsserierna, 1996-08-01. De första 500 observationerna används för att ta fram prognoserna för observation 501. Strategierna är definierade tidigare i uppsatsen. För att rekapitulera så beskriver vi dem kortfattat igen. Vi antar att vi kan placera i en småbolagsfond eller i en riskfri tillgång. Strategi S0 går ut på att satsa hela kapitalet i småbolagsfonden och stanna i placeringen genom hela simuleringen (Buy-Hold i tabell 4). Strategi S1 använder sig av VECM-prognoser, dvs den tar hänsyn till felkorrigeringstermen. Vi använder bara tecknet på prognosvärdet. När den prognostiserade differensen på index är positiv så går vi in med 100% i småbolagsfonden. När den prognostiserade differensen är negativ så placerar vi 100% av kapitalet i den riskfria tillgången. På samma sätt agerar vi för S2 och S3. Skillnaden är att vi använder VAR-prognoser respektive RW-prognoser för att estimerar $E(\Delta y_t)$.

De grå cellerna i tabellen anger det bästa värdet. Vi ser att prognosmodellerna får ett bättre utfall för alla prognosfönsterstorlekar upp till 1300. Härifrån så är buy-and-hold strategin mera fördelaktig. Det kan förklaras med att härifrån så stiger *Small Cap* konstant. Det är svårt att slå buy-and-hold strategi i en sådan konjunktur. Vi är primärt intresserade av de prognosstorlekarna som ligger runt 500. Dvs en storlek som motsvarar en typ av konjunkturtyp. Vi kan se att VECM-modellen presterar bättre än VAR-prognosmodellen för prognosfönster som ligger mellan 500 och 1400 sample. Det är ett tecken på att vårt ursprungliga gissning på prognosfönster passade modellen bra. RW-prognosmodellen som exkluderats i simuleringens utvärderingen presterar mycket bra. Vi kommer inte att gå vidare med analys av RW-resultatet.

5. SLUTSATSER

Vi behövde utveckla verktyg för att genomföra analysen och prognosmetoden. Teorin för prognostisering är omstridd. Beräkningarna ställer stora krav på bl a tillgång till information och låga transaktionskostnader. Vi har antagit att coutaget är noll och vi valde istället att minimera transaktionskostnaderna som härrör sig till arbetstid med beräkningar. Vi utvecklade makro som kan modifieras för användning inom program trading i realtid. Vi hade nytta av detta då vi gick en storlek på prognosintervall till att hantera 20 olika storlekar.

Internationella arbeten av bl a Arbeter (2003) påvisar prognoskraft hos size-komponenten. Vi utvecklade en bivariat korsautokorrelationsmodell i form av en VAR-modell. När vi kom på idén att använda den gemensamma trenden så hittade vi Engel-Grangers artiklar och vi utvecklade ytterligare en bivariat korsautokorrelationsmodell som inkluderade en felkorrigeringsterm. Prognosmodellerna utvärderade vi med ME, MAE och RMSE. Kriteriet för prognoskraft var att prognosmodellerna skulle prestera bättre än en RW-modell som utvecklats med samma differensstruktur som de övriga två modellerna. Prognosmodellerna bedömdes ha bättre *fit* än referensmodellen. Vi gjorde även en utvärdering med hjälp av simulering av handelsstrategier. Referensmodellen var här en passiv buy-and-hold strategi i *Small Cap*. Prognosmodellerna presterade bättre än referensmodellen med undantag för prognosfönsterstorlek på 1300 sample och uppåt. Detta kan ses som ett undantag på *Small Cap* gick konstant uppåt för denna period i kombination med att vi egentligen var intresserade av prognosfönster med storlek runt 500 sample. VECM kunde inte påvisa varken bättre *fit* eller bättre simuleringsresultat än VAR-prognosmodellen. Felkorrigerings termen bedöms som intressant vid vår estimering av småbolagsindex.

De övergripande bedömningskriterierna som definierades i slutet av kapitel 3 bedömdes uppfyllda. Prognosmodellerna hade bättre *fit* än Random Walk referensmodellen och de hade bättre ekonomisk signifikans än buy-and-hold referensmodellen. Eftersom de bivariata prognosmodellerna båda två bedömdes ha prognoskraft så kan vi dra slutsatsen

att det är sannolikt att marknaden inte är starkt effektiv och att det sker en gradvis informationsspridning från *Big Cap* till *Small Cap*.

Slutsatsen med avseende på vår arbetsmetod är att vi gjorde rätt i att välja makroberäkningar för att minimera tiden för beräkning. När vi gick över till 20 olika storlekar på prognosintervall så hade normala excelberäkningar inte varit hanterbara och dessutom mycket tidskrävande att hantera. Vi lärde oss mycket tankesättet vid prognostiseringsanalyser och vi har med oss erfarenheten till framtida analyser.

6. BILAGOR

BILAGA 1. Engel-Grangers 3-stegsmetod (Enders version)

BILAGA 2. Programscript för prognosmodell 1 – VAR

BILAGA 3. Programscript för prognosmodell 2 – VECM

BILAGA 4. Programscript som tar fram prognoser för en specifik samplingsperiod

7. ORDLISTA

Makro/Script	En namngiven sekvens av kommando som kan exekveras på en dator genom anrop.
Small Cap	Små bolag
Big Cap	Stora bolag med aktier som handlas frekvent
Korsautokorrelation	Ett samband mellan en tidsserie och tidigare värden av en annan tidsserie
Autokorrelation	Ett samband mellan en tidsserie och tidigare värden av samma tidsserie
Spurious Regression	Se definition i texten
Substitut	Ersättning. T ex så använder vi ett substitut för småbolag då en sådan aktie inte finns.
Program-Trading	Handel som utförs med automatik av en software.

8. REFERENSER

- Binner Jane M, Elger Thomas, Nilson Birger och Jonathan A Tepper (2004), "Tools for non-linear time series forecasting in economics- an empirical comparison of regime switching vector autoregressive models and recurrent neural networks, Application of artificial intelligence in Finance and Economics, Advances in Econometrics, Volume 19, page 71-92.
- Eugene F. Fama, Random walks in stock market prices, Financial Analysts Journal, 1995, Volume 51, Issue 1, Pages 75-81 (Reprint of 1965 article)
- Eugene F. Fama, Efficient Capital Markets: A Review of theory and Empirical Work, The Journal of Finance, Vol.25, No.2, 1970
- Eugene F. Fama, Market efficiency, long-term returns, and behavioral finance, Journal of Financial Economics, 1998, Vol.49, Issue 3, Pages 283-306
- Walter Enders. Applied Econometric Time Series 2nd Edition, John Wiley & Sons 2003, ISBN 0-471-23065-0
- Walter Enders., Rats Handbook for econometric time series, , John Wiley & Sons 1996, ISBN 0-471-14894-6
- Granger, C.W.J., Forecasting stock market prices: Lessons for forecasters, 1992, Volume 8, Issue 1, Pages 3-13
- Robert F. Engle; C. W. J. Granger , Co-integration and Error Correction: Representation, Estimation and Testing, Econometrica 55 (1987): 251-276.
- Louis Bachelier; G. U. Y. , Le Jeu, la Chance et le Hasard, Journal of the Royal Statistical Society, 1914, Volume 77, Issue 8, Pages 867-868
- Burton G. Malkiel, A Random Walk Down Wall Street: Completely Revised and Updated Eighth Edition, (April 2003) ISBN: 0393057828
- Hulbert, Mark , A schizophrenic walk down Wall Street., Forbes, Year 1996, Volume 158, Issue 10, Pages 318-319
- Chan Louis K C; Jegadeesh Narasimhan; Lakonishok Josef, Momentum strategies, The Journal of Finance, Year 1996, Volume 51, Issue 5, Pages 1681-1713
- K. Geert Rouwenhorst, International Momentum Strategies, Journal of Finance, Year 1998, Volume 53, Issue 1, Pages 267-284
- Narasimhan Jegadeesh; Sheridan Titman, Profitability of Momentum Strategies: An Evaluation of Alternative Explanations, Journal of Finance, Year 2001, Volume 56, Issue 2, Pages 699-720

- McCulloch W, Pitts W, (1943), “ A logical calculus of ideas immanent in nervous activity”,
Bulletin of mathematical biophysics, 5, page115-133
- Lee Charles M C; Swaminathan Bhaskaran, Price momentum and trading volume, The Journal
of Finance, Year 2000, Volume 55, Issue 5, Pages 2017-2069
- Hong Harrison; Lim Terence; Stein Jeremy C, Bad news travels slowly: Size, analyst coverage,
and the profitability of momentum strategies, The Journal of Finance, Year 2000,
Volume 55, Issue 1, Pages 265-295
- Moskowitz Tobias J; Grinblatt Mark, Do industries explain momentum? The Journal of
Finance, Year 1999, Volume 54, Issue 4, Pages 1249-1290
- Kahn Ronald N; Rudd Andrew, Does historical performance predict future performance?,
Financial Analysts Journal, Year 1995, Volume 51, Issue 6, Pages 43-53
- Mark Arbeter, Will the Big Caps back Down? Business Week,
2 Oktober 2006.
- Speidell, Lawrence S.; Graves, John D. The case for international small-cap stocks., Journal of
Investing, Euromoney Institutional Investor PLC, Year 1998, Volume 7, Pages 6-11
- Altay Erdinç, CROSS-AUTOCORRELATION BETWEEN SMALL AND LARGE CAP
PORTFOLIOS IN THE GERMAN AND TURKISH STOCK MARKETS, Journal of
Financial Management & Analysis, Year 2004, Volume 17, Pages 77-92
- Richardson Terry; Peterson David R , The cross-autocorrelation of size-based portfolio returns
is not an artifact of portfolio autocorrelation, The Journal of Financial Research, Year
1999, Volume 22, Issue 1, Pages 1-13
- C. W. J. Granger; T. H. Lee, Investigation of Production, Sales and Inventory Relationships
Using Multicointegration and Non-Symmetric Error Correction Models, Journal of
Applied Econometrics, Year 1989, Volume 4, Pages S145-S159

<http://www.carnegie.se>

<http://www.omxgroup.com>

BILAGA 1. Engel-Grangers 3-stegsmetod (Enders version)

Denna version ligger mycket nära den som finns i Enders bok ”*Rats Handbook for econometric time series, 1996, John Wiley & sons*”, sidan 156-173. Anta att vi har två variabler y och x och att de är integrerade av första ordningen. Vi vill ta reda på om det finns ett jämviktförhållande mellan de två. Engel och Granger (1987) föreslår ett test för att ta reda på om de två variablerna som är $I(0)$ är kointegrerade av ordning $CI(1,1)$.

Steg 1

Testa om de två variablerna är integrerade av samma ordning. Gör det med ett Dickey-Fuller test. Om båda tidsserierna är stationära så finns det ingen anledning att fortsätta för då kan man använda vanlig tidsserieanalys. Om variablerna är integrerade av olika ordning så är de inte kointegrerade så som tänkt. Det kan hända att en variabel som är $I(1)$ är kointegrerad med en som är $I(2)$ men då krävs att det finns flera variabler och att grupperna är kointegrerade och då talar vi om multicointegration. Detta kommer vi att exkludera från vår analys. Med andra ord, om y och x är $I(1)$ så kan vi fortsätta till steg 2.

Steg 2

Nu ska vi estimerade det långsiktiga jämvikten.

$$y_t = a + bx_t + e_t \quad (\text{Engel-Granger Steg 2})$$

Om variablerna är kointegrerade så får vi en ”superkonsistent” estimator på kointegrationsparametrarna. Stock (1987) bevisar att koefficienterna konvergerar snabbare än i OLS modellen som använder stationära variabler. Tittar vi på en scatterplot för de två variablerna så ser vi att effekten av trenden dominerar den stationära komponenten. Där ska finnas ett starkt linjärt samband. Vi använder tidsserier för residualerna. Dessa är de estimerade värdena av avvikelserna från jämvikt. Om dessa är $I(0)$ så är y_t och x_t är $CI(1,1)$. En speciell version av Dickey Fuller testet används eftersom vi ska testa skattade värden av e och inte de verkliga. Engel och Granger tillhandahåller en speciell uppsättning av kritiska värden som vi använder.

Steg 3

Residualerna från regressionen används för att skatta error correction modellen.

$$\Delta y_t = a_{10} + a_y(y_{t-1} - bx_{t-1}) + a_{11}\Delta y_{t-1} + a_{12}\Delta x_{t-1} + e_{yt}$$

$$\Delta x_t = a_{20} + a_x(y_{t-1} - bx_{t-1}) + a_{21}\Delta y_{t-1} + a_{22}\Delta x_{t-1} + e_{xt}$$

Engel och Granger (1987) föreslår en metod för att förenkla modellen. \hat{e}_{t-1} är avvikelserna från jämvikt period $t-1$. Därför kan vi använda de sparade residualerna från steg 2 som ett instrument för uttrycket $y_{t-1} - bx_{t-1}$.

$$\Delta y_t = a_{10} + a_y \hat{e}_{t-1} + a_{11}\Delta y_{t-1} + a_{12}\Delta x_{t-1} + e_{yt} \quad (\text{Engel-Granger Steg 3})$$

samt motsvarande ekvation för x .

Det är även denna vektor som man använder för att prognostisera med hjälp av ECM.

Utöver felkorrektions termen så har vi en VAR modell i differensform. Det som man kan göra med en VAR modell kan man göra med denna.

Steg 4

Bedöm modellens kvalitet.

Speed of adjustment är a_y och a_x . Om de är noll så har vi en VAR. Om alla utom en variable är svagt exogen så kan man vilja använda en annan teknik som kallas general-to-specific. Vi går bara djupare in i detta steg om vi får problemet i vår analys.

BILAGA 2. Programscript för prognosmodell 1 - VAR

VAR-prognoser beräknas. Prognosfönstrets storlekt hämtas från excelarket. Prognosen beräknas i ett steg.
Prognoserna estimeras med *prognosmodell 1 (VAR)*. *Proceduren upprepas tills att man når tidsseriernas slut.*

```
Sub VAR_calc()
' Version 1.0.0 Build 00002 (Release 2007-05-01)
' Utvecklare: Måns Näsman
' Språk: Visual Basic 6.3

' Detta program hämtar och skriver data i excelbladet "02".
' Programmet skattar och beräknar VAR-prognoser för t ex observationerna 501-2297.

Dim AntalObs As Long ' Prognostiseringsfönster
AntalObs = Sheets("00").Range("q9:q9").Value
Set count1 = Sheets("00").Range("s12:s12")
count1.Value = ""

Dim sy As String ' Y (t ex 498 st obs.)
Dim sx As String ' X (t ex 498 st obs med vektor om 2 st oberoende variabler)
Dim ry As Range ' Y (t ex 498 st)
Dim rx As Range ' X (t ex 498 st obs med vektor om 2 st oberoende variabler)
Dim rk As Range ' Koefficientvektor (a, b, c)

' Kolumnernas innehåll.
' B = Y
' C = X
' E-g = Koefficienter. Rad 11.
' L = DY
' M = DY(-1)
' N = DX(-1)
' Q = E(DY)

For i = 0 To (2296 - AntalObs)

    j = i
    count1.Value = i

    ' Genomför estimering (baseras på 498 observationer)
    sy = "l" & (6 + j) & ":l" & (AntalObs + 3 + i)
    sx = "m" & (6 + j) & ":n" & (AntalObs + 3 + i)
    Set ry = Sheets("02").Range(sy)
    Set rx = Sheets("02").Range(sx)
    Set rk = Sheets("02").Range("e11:g11")
    rk = Application.WorksheetFunction.LinEst(ry, rx)

    ' Beräkna prognosvärde
```

```
x1 = Sheets("02").Cells(AntalObs + 3 + i, 13).Value
x2 = Sheets("02").Cells(AntalObs + 3 + i, 14).Value
a = Sheets("02").Cells(11, 7).Value
b = Sheets("02").Cells(11, 6).Value
c = Sheets("02").Cells(11, 5).Value

edy = a + b * x1 + c * x2
Sheets("02").Cells(AntalObs + 4 + i, 17).Value = edy
```

Next

End Sub

BILAGA 3. Programscript för prognosmodell 2 - VECM

VECM prognoser beräknas. Prognosfönstrets storlek hämtas från excelarket. Den första prognosen beräknas i två steg. Först estimeras residualtermerna som motsvarar *Engel-Granger steg 2* i bilaga 1. Residualerna används för att göra en ny estimering med ett nytt fönster. Prognoserna estimeras med *prognosmodell 2 (VECM)*. *Proceduren upprepas tills att man når tidsserierna slut.*

```
Sub VECM_calc()
' Version 1.0.0 Build 00002 (Release 2007-05-01)
' Utvecklare: Måns Näsman
' Språk: Visual Basic 6.3

' Detta program hämtar och skriver data i excelbladet "01".
' Programmet skattar och beräknar VECM-prognoser för t ex observationerna 501-2297.

Dim AntalObs As Long ' Prognostiseringsfönster
AntalObs = Sheets("00").Range("q9:q9").Value
Set count1 = Sheets("00").Range("s11:s11")
count1.Value = ""

Dim sy1 As String ' Y (t ex 500 st obs.)
Dim sx1 As String ' X (t ex 500 st obs.)
Dim ry1 As Range ' Y (t ex 500 st obs.)
Dim rx1 As Range ' X (t ex 500 st obs.)
Dim rk1 As Range ' Koefficientvektor steg 1 (a, b)

Dim sy2 As String ' Y (t ex 500-2=498 st obs.)
Dim sx2 As String ' X (t ex 500-2=498 st obs med vektor om 3 st oberoende variabler)
Dim ry2 As Range ' Y (t ex 500-2=498 st)
Dim rx2 As Range ' X (t ex 500-2=498 st obs med vektor om 3 st oberoende variabler)
Dim rk2 As Range ' Koefficientvektor steg 2 (a, b, c, d)

' Kolumnernas innehåll.
' B = Y
' C = X
' E-H = Koefficienter. Steg 1 på rad 7. Steg 2 på rad 11.
' J = Feltermen steg 1
' L = DY
' M = e(-1)
' N = DY(-1)
' O = DX(-1)
' Q = E(DY)

For i = 0 To (2296 - AntalObs)

    j = i
    count1.Value = i
```



```

' Genomför Steg 1 estimering (baseras på t ex 500 observationer)
sy1 = "b" & (4 + j) & ":b" & (AntalObs + 3 + i)
sx1 = "c" & (4 + j) & ":c" & (AntalObs + 3 + i)
Set ry1 = Sheets("01").Range(sy1)
Set rx1 = Sheets("01").Range(sx1)
Set rk1 = Sheets("01").Range("e7:f7")
rk1 = Application.WorksheetFunction.LinEst(ry1, rx1)

' Genomför Steg 2 estimering (baseras på t ex 500-2=498 observationer)
sy2 = "l" & (6 + j) & ":l" & (AntalObs + 3 + i)
sx2 = "m" & (6 + j) & ":o" & (AntalObs + 3 + i)
Set ry2 = Sheets("01").Range(sy2)
Set rx2 = Sheets("01").Range(sx2)
Set rk2 = Sheets("01").Range("e11:h11")
rk2 = Application.WorksheetFunction.LinEst(ry2, rx2)

' Beräkna prognosvärde
x1 = Sheets("01").Cells(AntalObs + 3 + i, 13).Value
x2 = Sheets("01").Cells(AntalObs + 3 + i, 14).Value
x3 = Sheets("01").Cells(AntalObs + 3 + i, 15).Value
a = Sheets("01").Cells(11, 8).Value
b = Sheets("01").Cells(11, 7).Value
c = Sheets("01").Cells(11, 6).Value
d = Sheets("01").Cells(11, 5).Value

edy = a + b * x1 + c * x2 + d * x3
Sheets("01").Cells(AntalObs + 4 + i, 17).Value = edy

Next

End Sub

```

BILAGA 4. Programscript som tar fram prognoser för en specifik samplingsperiod

I excelbladet anger man prognosfönstrets storlek. Makrot nedan anropas i samband med att man klickar på en kommandoknapp
Makrot nollställer kalkylarkens prognoscolumn och anropar sedan prognosberäkningarna för VAR och VECM.
Resultatet blir tidsserier för prognoserna. Prognoserna används för en automatiskt prognosutvärdering som sammanställs
I excelbladet.

```
Sub Total_calc()
```

```
    Set ry1 = Sheets("01").Range("q5:q2300")  
    ry1.Value = ""  
    Set ry1 = Sheets("02").Range("q5:q2300")  
    ry1.Value = ""  
    'Set ry1 = Sheets("03").Range("q5:q2300")  
    'ry1.Value = ""  
    Set ry1 = Sheets("03").Range("p5:p2300")  
    ry1.Value = ""  
    a = Sheets("00").Range("q9:q9").Value + 4  
    Set ry1 = Sheets("03").Range("p" & a & ":p2300")  
    ry1.Value = "1"
```

```
    VECM_calc
```

```
    VAR_calc
```

```
End Sub
```