



Centrum för  
bank och finans

# Framtagning av komposita prisindex

Han-Suck Song och Mats Wilhelmsson

(KAFI AB och Centrum för bank och finans, KTH)

på uppdrag av Valueguard AB  
finansierat av Vinnova

Stockholm

2010-05-25

## 1. Introduktion

För närvarande (våren 2010) finns det i Sverige tre stycken NasdaqOMX Valueguard-KTH bostadsrättsindex för regionerna Stockholm, Göteborg och Malmö. Dessa går under benämningen HOX och uppdateras månatligen. Det finns skäl för att man skall ta fram ett nationellt index som är ett kompositindex bestående av de tre regionala indexserierna plus eventuellt några fler regioner.

Syftet med följande rapport är att analysera följande två viktiga förutsättningar vid framtagandet av ett nationellt bostadsrättsprisindex:

- Vilken viktningsmetod som ska användas för att beräkna de ingående regionernas vikt in det nationella indexet.
- Hur man kan fastställa vilka regioner, och därmed antalet regioner, som bör ingå i det nationella indexet.

Valet av viktningsmetod baseras på Monte Carlo simuleringar. För att fastställa antalet regioner, det vill säga vilka regioner som bör ingå i det nationella indexet, utförs ekonometriska analyser.

Uppsatsen är disponerad enligt följande: i nästkommande avsnitt diskuteras de överväganden som man måste hantera vid beräkningen av ett kompositindex, det vill säga urvalet av regioner och viktningen av dessa. I avsnitt 3 presenteras vikter som används i olika delar av världen. I avsnitt 4 presenteras Monte Carlo simuleringar som syftar till att analysera hur olika typer av viktningsförfaranden kan påverka kompositindexet. Tolkningar av resultatet från simuleringarna presenteras i avsnitt 5. Frågan om hur många regioner som ska ingå i det nationella indexet diskuteras i avsnitt 6. Slutsatser och rekommendationer presenteras i avsnitt 7.

## 2. Överväganden

Att ta fram ett månatligt kompositindex eller nationellt index utifrån ett antal regionala prisindexserier kan göras på ett antal olika sätt och hur man bör gå tillväga är inte ett trivialt problem. Först måste de regioner som ska vara med i indexet fastställas. Därefter behöver varje ingående region tilldelas en vikt i det nationella indexet.

Valet av viktningsystem bygger på olika överväganden. Å den ena sidan bör vikterna medföra att stora marknader med många transaktioner och höga prisnivåer ska ha en stor påverkan på det komposita indexet. Å andra sidan ska inte omviktningar kunna leda till att det nationella indexet uppvisar stora svängningar som beror på omviktningar.

Anta exempelvis att vi har två jämstora marknader både när det gäller prisnivåer och antal fastigheter. Här vill vi att vikten skall vara 0.5 för respektive ort, det vill säga, de har en lika stor påverkan på det nationella indexet. Anta nu att omsättningen (mätt i antalet transaktioner) minskar i en region (på grund av en strejk bland vissa fastighetsmäklarna).

Om vikterna var baserade på omsättningen varje månad skulle det innebära att regionen med den lägre omsättningen viktas ned. Detta skulle i sin tur kunna medföra att det nationella indexet inte förändras på grund av prisförändringar utan på grund av att antalet transaktioner förändras. I tabellen nedan åskådliggörs vad som skulle kunna hända om vikterna förändras kortsiktigt.

Tabell 1. Exempel på om vikterna ändras.

Månad	Index		Urspr. vikt		Nationellt index	Ny vikt		Nationellt index	Delta	
	Region 1	Region 2	Region 1	Region 2	urspr. Vikt	Region 1	Region 2	ny vikt	urspr	ny
1	100	100	0,5	0,5	100	0,5	0,5	100		
2	102	100	0,5	0,5	101	0,5	0,5	101	1,0%	1,0%
3	103	100	0,5	0,5	102	0,5	0,5	102	0,5%	0,5%
4	105	98	0,5	0,5	102	0,5	0,5	102	0,0%	0,0%
5	105	101	0,5	0,5	103	0,5	0,5	103	1,5%	1,5%
6	106	105	0,5	0,5	106	0,5	0,5	106	2,4%	2,4%
7	108	105	0,5	0,5	107	0,5	0,5	107	0,9%	0,9%
8	108	104	0,5	0,5	106	0,5	0,5	106	-0,5%	-0,5%
9	109	106	0,5	0,5	108	0,5	0,5	108	1,4%	1,4%
10	110	105	0,5	0,5	108	0,25	0,75	106	0,0%	-1,2%
11	112	105	0,5	0,5	109	0,25	0,75	107	0,9%	0,5%
12	115	105	0,5	0,5	110	0,5	0,5	110	1,4%	3,0%
Standardavvikelse									0,8%	1,2%

Nedgången i antalet transaktioner månad 10 och 11 i region 1 innebär att vikten avseende region 1 minskar och därmed ökar vikten för region 2. Vilka effekter får detta? I det här fallet så visade att det ursprungliga kompositindexet på en svag uppgång beroende på att region 1 har haft en stark prisuppgång medan region 2 uppvisar en svagare uppgång. Den nya viktningen innebär att prisuppgången förbytt mot en nedgång månad 10 och en svag uppgång månad 11. Det blir ingen skillnad under perioden 1-9. Månad 12 (där viktningen inte har påverkats) kommer dock att påverkas och nu redovisa på en stark uppgång. De nya vikterna har inneburit att den nya indexserien har en högre volatilitet som helt har förorsakats av omviktningen.

Den andra centrala frågan, nämligen hur många regioner som bör ingå i det nationella indexet, är inte heller ett trivialt problem. Å ena sidan är det bättre ju fler av landets regioner som ingår i ett nationellt index. Å andra sidan kan det uppstå beräkningstekniska problem om regioner med fåtal transaktioner ingår. Att valet av antalet regioner kan vara ett svårt problem framgår av hur S&P/Case-Shiller boprisindex har ändrat antalet ingående regioner: idag ingår 20 regioner i USA S&P/Case-Shiller nationella boprisindex. Tidigare ingick endast 10 regioner. De regioner eller orter som man inte har täckning för ingår inte i beräkningarna. Vår utgångspunkt är att hela Sverige skall ingå när det nationella kompositindexet beräknas. Ett första test kan utgå ifrån om det räcker med tre regionala storstadsindex plus ett index som täcker in resten av Sverige.

### 3. Vilka typer av vikter används idag?

Den typ av vikt som illustrerades ovan (vikter som ändras varje månad med hänsyn till förändringar i fördelningen av antalet transaktioner i de ingående regionerna) används idag men är kanske inte den vanligaste typen. Gör man en internationell utblick kan man konstatera att det finns en relativt stor bredd när det gäller vilken typ av viktning man använder sig av.

## Använda viktningsmetoder i världen

Som nämnts ovan är det några typer av viktningsmetoder som används. Följande tabell baseras på bland annat Lum (2004) samt Levy och Post (2005).

Tabell 2. Viktningsmetoder

Land	Viktningsmetod	Uppdateras
Singapore	Värdebaserat, transaktioner	"Moving average" 12 kvartal
Hong-Kong	Värdebaserat, transaktioner	Varje månad
Kanada 1	Värdebaserat, transaktioner	"Moving average" 3 år, ändras årligen
Kanada 2	Värdebaserat, bestånd	2006 "Canada Census"
UK	Värdebaserat, transaktioner	Varje kvartal
USA	Värdebaserat, bestånd	Vid ny "Census", laggad 2-3 år.

Tabell 2 visar att viktningar som baseras på antalet transaktioner är fler än de som baseras på beståndet. Även hur ofta som uppdateringar av vikter görs skiljer sig åt. Vissa index uppdateras vikter varje månad medan andra gör det mycket mer sällan. En variant är att vikterna uppdateras varje månad, men att de baseras på det genomsnittliga antalet transaktionerna den senaste tolv månadersperioden. Ett motiv till detta är att undvika att stora månatliga variationer i antalet sålda objekt skall påverka viktningen och därmed indexberäkningen. De olika viktningsmetoderna som används idag tar normalt även hänsyn till genomsnittliga bostadsvärden i de ingående regionerna. Med andra ord, indexen är antingen beståndsbaserade värdevikter eller transaktionsbaserade värdevikter enligt följande.

- (1) Beståndsbaserade värdevikter. Det är vikter som utgår ifrån beståndet av fastigheter/lägenheter i en viss region. Antalet bostäder multipliceras med det genomsnittliga värdet. Dessa vikter kan beräknas:
  - a. Initialt – man utgår från bastidpunkt. Årliga vikter – innebär att man beräknar om vikterna en gång om året baserat på beståndet och värdena de senaste 12 månaderna.
  - b. Månadsvikt – vikterna ändras varje månad beroende på om beståndet och/eller värdena har förändrats.
  - c. "Moving average" – man uppdaterar vikterna varje månad men de baseras på ett 12 månader rullande medelvärde. Man kan även använda sig av exempelvis 24 eller 36 månaders rullande medelvärde.
  - d. Beståndsvikter – här låter man endast beståndet påverka vikterna och inte värdet.
- (2) Transaktionsbaserade värdevikter. Det är vikter som istället för beståndet baseras på antalet transaktioner. Fördelen kan vara att mindre information om marknaden behövs och nackdelen är att omsättningen kan skilja sig åt mellan olika regioner och över tiden vilket kan få en inverkan på det komposita indextalet. Även de transaktionsbaserade värdevikterna beräknas enligt ovan, det vill säga initialt, årliga, månatliga, samt "moving average".

Ett tredje sätt som kan användas är att man beräknar det nationella indexet direkt utifrån en hedonisk prisekvation, det vill säga man beräknar inte flera olika regionala indexserier som man sedan viktar ihop. Detta kan göras så att man tillåter regionsspecifika implicita priser eller inte. Om man tillåter regionsspecifika implicita priser så kommer "viktningen" att likna den transaktionsbaserade värdeviktningen där man har olika vikter varje månad.

### ***"S&P/Case-Shiller U.S. National Home Price Indexes"***

"S&P/Case-Shiller U.S. National Home Price Indexes" - det nationella bostadsprisindexet i USA - består inte av samtliga regioner i landet utan endast av de 20 största storstadsregionerna. Tidigare ingick endast 10 regioner. Anledningen är att man inte har täckning i alla stater. I tretton stater har man ingen täckning alls och i nio stater täcker man in mindre än 75 procent av det totala bostadsvärdet (se Rapport, 2007).

Viktningen av de 10 alternativt 20 regionala indexerna har skett med en beståndsbaserad värdeviktning. Enligt Rappaport (2007) har man utgått ifrån beståndet år 2000 och beräknat värdet på dessa med hjälp av de (genomsnittliga) bostadspriserna. Det innebär att deras viktningss metod liknar metoden 1a ovan. Man kan också likna deras värdering vid den viktning man gör för att konstruera aktieindexet S&P500 (eller för den delen OMX30, se nedan).

Att man använder sig av år 2000 som bas framgår även av "S&P/Case-Shiller Home Price Indices – Index Methodology" daterat november 2009. Där skriver man "Until new Census estimates of the aggregate value of single-family housing units become available, the 2000 base period estimates of aggregate value will be used for calculating updates to the U.S. national home price index."

### ***Aktieindex***

Om man istället sneglar på den viktning som NasdaqOMX gör när de beräknar OMXS30 kan man konstatera att den ligger nära den beståndsbaserade värdeviktningen presenterad ovan.

Enligt "Regelverket för OMX derivatives markets" är konstruktionen av OMXS30 gjord med utgångspunkten att för "ett begränsat antal aktier skapa ett index som så långt som möjligt utvecklas i överensstämmelse med aktierna noterade på Børsen". I OMXS30 ingår 30 aktier. De som ingår har den högsta omsättningen mätt i svenska kronor på Børsen under en viss mätperiod. Mätperioden är den sexmånadersperiod som börjar sju månader innan ett kalenderhalvår. OMXS30 viktningen ändras varje halvår.

Om en aktie som är inkluderad i OMXS30 inte tillhör de 45 mest omsatta aktierna så skall aktien ersättas. Om en aktie som inte ingår i OMXS30 tillhör de 15 mest omsatta skall den inkluderas i index och ersättas den som har lägst omsättning. Dessutom finns regler om få avslut, offentligt uppköpserbjudande etc. som påverkar om aktien ingår eller inte ingår.

Beräkningen av indexvärdet en viss given tidpunkt får man fram genom att beräkna summan av börsvärdet för en aktie (kurs\*antal aktier) dividerat med börsvärdet för samtliga aktier som ingår i index. Beräkningen korrigeras för exempelvis emissioner och konverteringar. Se Levy and Post (2005) för en djupare diskussion om olika indexviktningar.

## ***Slutsatser angående skillnader mellan HOX, OMXS30 och OMX Stockholm PI***

Valet av just 30 aktier har skett *ad hoc*. Man hade lika väl kunnat välja 20 aktier eller något annat. Vilka 30 bolag som ingår är dock inte *ad hoc* då den baseras på hur många som har omsatts under en viss period (beräknas i kronor) och uppdateras varje halvår.

Viktningen har gjorts med utgångspunkt utifrån storlek på aktie och i det här fallet kursen multiplicerat med antalet aktier enligt gällande aktielista. Viktningen görs om varje tidsenhet så man skulle kunna karakterisera OMXS30 som ett beståndsbaserat värdeindex där viktningen är lika med 1b ovan. Det finns ingen diskussion om varför man har valt att uppdatera ingående aktier varje halvår och inte varje månad eller varje dag. Valet av halvår ser därför ut att vara relativt *ad hoc*.

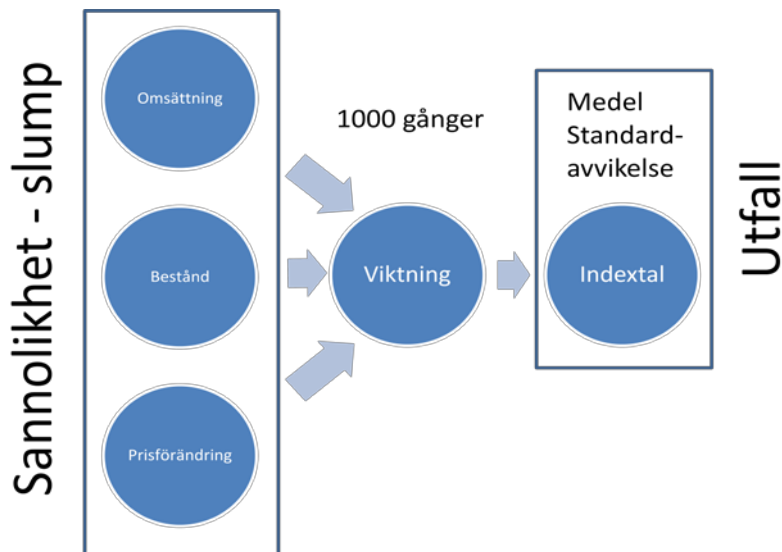
Det komposita HOX-indexet har en större likhet med OMX Stockholm PI än OMXS30 att exempelvis jämföras med S&P/Case-Shiller index som jämförs sig själva med S&P500. Vår avsikt är däremot att beräkna ett HOX-index avseende bostadsrättspriser i Sverige ("HOX BR Sverigeindex"). Det innebär således att samtliga delar av Sverige bör vara inkluderat i indexet, men till skillnad från exempelvis OMX Stockholm PI där man kan beräkna börsvärdet varje tidpunkt är det inte möjligt att beräkna HOXens motsvarande månadsindex för samtliga kommuner i Sverige. Det innebär att beräkningen av "HOX BR Sverigeindex" måste bestämma vissa principer angående hur samtliga delar av Sverige skall kunna ingå. Ett alternativ är naturligtvis att beräkna exempelvis "HOX BR5", där endast de 5 största storstadsregionerna ingår.

### **4. Simulering av effekten av att använda sig av de olika viktningarna.**

Monte Carlo metoden är en metodik där slumpmässig simulering används för att skatta parametrar. Monte Carlo metodiken används ofta när det inte finns någon analytisk lösning på ett problem. Här används metodiken för att skatta prisindexparametrar där olika typer av viktningar används och där man låter pristrend, omsättning och bestånd variera. Anledningen till att vi inte kan använda faktisk data är att det inte existerar data under olika ekonomiska marknadsförhållanden.

I nedanstående beskrivs simuleringen schematiskt. Utfallen i beräkningarna genererar indextal avseende ett kompositindex. Varje utfall beror på tre följande faktorer som alla påverkas av slumpen: prisutvecklingen, omsättningen, och beståndet. Samtliga dessa tre storheter kan påverka viktningen som sedan kan påverka de enskilda indextalen. Genom att simulera fram flera utfall (här 1000 gånger) är det möjligt att åskådliggöra effekterna av olika viktningmetoder som presenterats ovan.

Figur 1. Schematisk beskrivning av experimentet.



### Viktning

Det som framförallt kommer att utvärderas här är skillnaden mellan viktningssmetod 1 och 2. Eftersom 3:an är en variant av 2:a, kommer den inte att utvärderas. Metod 4 är inte aktuell då vi inte analyserar nyproduktion. Samtliga viktningssmetoder är baserade på värde.

- V1: Bestånd\*värde år t-12
  - Viktningen bestäms av värdet på fastigheterna i respektive region. Värde och bestånd avser året innan indexberäkningarna påbörjas. Uppdateras ej.

$$V1_{R1,t} = (\bar{P}_{R1,t-12} * \dot{S}_{R1,t-12}) / [(\bar{P}_{R1,t-12} * \dot{S}_{R1,t-12}) + (\bar{P}_{R2,t} * \dot{S}_{R1,t-12})]$$

- Där R1 är region 1, R2 region 2 och t är lika med månad. Vidare är P lika med pris och  $\bar{P}$  lika med genomsnittligt pris. Beståndet är lika med S och  $\dot{S}$  innebär att det är konstant över tiden.
- V2: Bestånd\*värde år t-12, t, t+12, t+24, t+36, t+48
  - Viktningen bestäms av värdet på fastigheterna i respektive region. Värde och bestånd avser året innan varje indexberäkning. Uppdateras varje år.

$$V2_{R1,t} = (\bar{P}_{R1,t-12} * S_{R1,t-12}) / [(\bar{P}_{R1,t-12} * S_{R1,t-12}) + (\bar{P}_{R2,t-12} * S_{R1,t-12})]$$

- V3: Bestånd\*värde varje månad
  - Viktningen bestäms av värdet på fastigheterna i respektive region. Värde och bestånd beräknas och uppdateras varje månad.

$$V3_{R1,t} = (\bar{P}_{R1,t} * S_{R1,t}) / [(\bar{P}_{R1,t} * S_{R1,t}) + (\bar{P}_{R2,t} * S_{R1,t})]$$

- V4: Bestånd\*värde varje månad, men baserad på medelvärdet de 12 föregående månaderna
  - Viktningen bestäms av värdet på fastigheterna i respektive region. Värde och bestånd beräknas och uppdateras varje månad. Medelvärde över 12 månader.

$$V4_{R1,t} = (\bar{P}_{R1,t-12} * \bar{S}_{R1,t-12}) / [(\bar{P}_{R1,t-12} * \bar{S}_{R1,t-12}) + (\bar{P}_{R2,t-12} * \bar{S}_{R1,t-12})]$$

- Där  $\bar{S}$  är det genomsnittliga beståndet de senaste 12 månaderna.
- V5: Transaktioner\*värde år t-12
  - Se V1 men transaktioner istället för bestånd.
- V6: Transaktioner\*värde år t-12, t, t+12, t+24, t+36, t+48
  - Se V2 men transaktioner istället för bestånd.
- V7: Transaktioner\*värde varje månad
  - Se V3 men transaktioner istället för bestånd.
- V8: Transaktioner\*värde varje månad men baserad på medelvärdet de 12 föregående månaderna
  - Se V4 men transaktioner istället för bestånd.

Detta innebär att åtta olika komposita indexserier kommer att beräknas. Den viktningssmetod som kan anses vara optimal är V3, det vill säga viktningar som beräknas varje månad och som är beräknas utifrån beståndsbaserade värden. Viktningarna kommer att vara identiska om beståndet, antalet transaktioner och värdet är konstanta under perioden. Så är dock inte fallet, utan 16 olika simuleringar kommer att utföras enligt tabellen nedan. Varje simulering utförs 1000 gånger och genomsnittliga värden avseende medelvärden och standardavvikelse för enskilda månadsskattningar kommer att beräknas. Avsikten är att skatta känsligheten i parameterskattningarnas avseende förändringar i prisutvecklingstakt samt volatiliteten i prisutvecklingen, antalet omsatta transaktioner, samt om regionen växer eller ej (baserat på förändringar i beståndet).

### ***Simuleringsantaganden***

De simuleringar som kommer att göras avser ett kompositindex uppbyggt av två regionala indexserier. Prisutvecklingen över femårsperioden som indexet konstrueras antas följa en "random walk with a drift", det vill säga

$$PI_t = \alpha + \beta PI_{t-1} + \varepsilon_t$$

där  $PI$  är lika med prisindex och  $t$  är lika med tidsperiod (månad i detta fall). Feltermen  $\varepsilon$  är vanligtvis normalfördelad. Här använder vi oss av en enklare rektangelfördelning (uniform distribution). Det gör att vi har "för höga" sannolikheter att få extrema värden (det vill säga tunga svansar). På så sätt skapar vi ett "worse case" scenario. Prisutvecklingen bestäms av den autoregressiva delen plus en drift.

Beståndet i regionerna är känt liksom antalet transaktioner per månad tillika genomsnittliga värden av bostäderna i regionerna. Omsättningen kommer dock att slumpmässigt variera mellan de två regionerna. Omsättningen bestäms i enlighet med följande formel:

$$O_t = S_t(0,5 + \rho)$$

Där  $O$  avser omsättning i antal transaktioner och  $\rho$  avser slump termen. Den har en uniform fördelning där varje utfall inom ett intervall kan realiseras med lika sannolikhet. I ursprungsmodellen kommer omsättningen att vara lika med 5 procent av beståndet. I England avseende småhus så ligger omsättningen på ungefär 3-7 procent av beståndet varje år. Ett antagande om att 5 procent per månad är såldes ett väldigt högt antagande. Vidare kommer beståndet kommer att vara konstant över tiden i en av regionerna men öka trendmässigt i den andra regionen. Det innebär att beståndet bestäms av följande formel.

$$S_t = S_{t-1}(1 + (\gamma/12))$$

Det innebär att beståndet  $S$  i period  $t$  bestäms av beståndet i föregående period plus en trendmässig ökning om  $\gamma$  procent. De komposita nationella indexserierna kommer att vikta ihop i enlighet med ovanstående viktningsmetoder och redovisas i detalj nedan.

Tabell 2. Antaganden i respektive simulering.

	Pristrend ( $\alpha$ )		Prisslump ( $\epsilon$ )	Omsättningsslump ( $\rho$ )		Beståndstrend ( $\gamma$ )	
	Region 1	Region 2		Region 1	Region 2	Region 1	Region 2
Simulering 1	1	-1	+/- 2	+/- 1	+/- 1	1	0
Simulering 2	1	-1	+/- 5	+/- 1	+/- 1	1	0
Simulering 3	1	-1	+/- 2	+/- 3	+/- 1	1	0
Simulering 4	1	-1	+/- 5	+/- 3	+/- 1	1	0
Simulering 5	1	-1	+/- 2	+/- 1	+/- 1	2	0
Simulering 6	1	-1	+/- 5	+/- 1	+/- 1	2	0
Simulering 7	1	-1	+/- 2	+/- 3	+/- 1	2	0
Simulering 8	1	-1	+/- 5	+/- 3	+/- 1	2	0
Simulering 9	3	-1	+/- 2	+/- 1	+/- 1	1	0
Simulering 10	3	-1	+/- 5	+/- 1	+/- 1	1	0
Simulering 11	3	-1	+/- 2	+/- 3	+/- 1	1	0
Simulering 12	3	-1	+/- 5	+/- 3	+/- 1	1	0
Simulering 13	3	-1	+/- 2	+/- 1	+/- 1	2	0
Simulering 14	3	-1	+/- 5	+/- 1	+/- 1	2	0
Simulering 15	3	-1	+/- 2	+/- 3	+/- 1	2	0
Simulering 16	3	-1	+/- 5	+/- 3	+/- 1	2	0

Pristrenden anges som en procentuell uppgång per månad. I Stockholm har den genomsnittliga ökningen per månad varit 0.88 procent mellan 2005 och 2009 för bostadsrätter. Det innebär att ett antagande om 3 procent uppgång per månad utgör en mycket kraftfull ökning. Volatiliteten mäts som standardavvikelse och standardavvikelsen per månad i Stockholm har varit 3,89 under samma period som ovan. Vi använder oss av 2 respektive 5 procent per månad i simuleringarna, det vill säga:

$$\epsilon \sim U(-2,2) \text{ eller } U(-5,5)$$

I grundmodellen antas att beståndet är konstant i en region och plus 1 procent per år i den andra regionen. Valet av 1 procent motiveras av att det motsvarar ungefär deprecieringsfaktorn

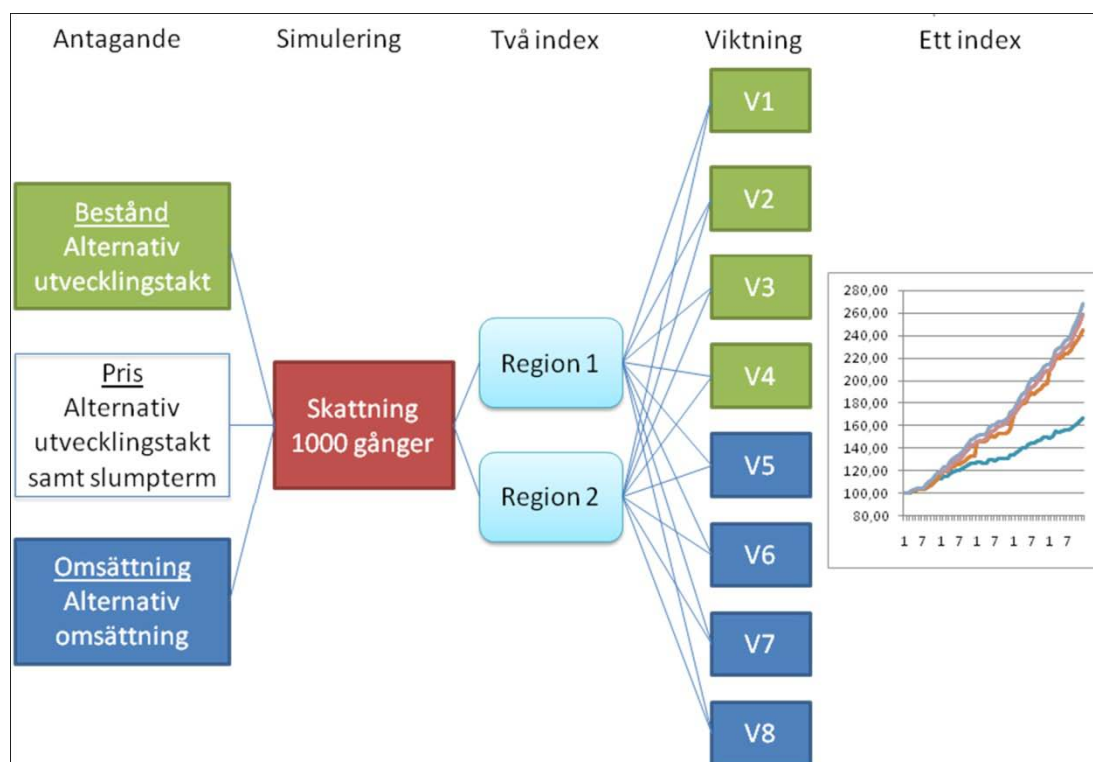
på en bostadsmarknad i jämvikt (se Wilhelmsson, 2008). Vi har också använt oss av den dubbla deprecieringsfaktorn, det vill säga 2 procent per år, för att analysera en kraftigt växande region.

Omsättning antas ligga på 5 procent av beståndet per månad med en slumpterm på plus/minus 1 procent (med lika fördelning). Maximal omsättning kan därmed vara 6 procent och minimalt 4 procent i grundmodellen. Slumptermen kommer att ökas till plus/minus 3 procent i en av regionerna. Det innebär att omsättningen kan variera från 2 till 8 procent av beståndet. Det innebär att

$$\rho \sim U(-1,1) \text{ eller } U(-3,3)$$

I figur 2 redogörs för Monte Carlo simuleringen. Bestånd, pris och omsättning beror delvis på slumpen. Indextal skattas för de två regionerna. Detta görs 1000 gånger. Därefter viktas de regionala ihop till ett index. Denna viktning kan ske på åtta olika sätt. Antagande om beståndet kommer att påverka viktningarna V1-V4. Antaganden om pris påverkar alla viktningssätt och antaganden om omsättningen kommer att påverka viktningssätt V5-V8.

Figur 2. Monte Carlo simulering.



### Simuleringsresultat

Kompositindexet (KI) beräknas i enlighet med nedanstående formel (V1-viktningen exemplifierar):

$$KI_t = (PI_{R1,t} * V1_{R1,t}) + (PI_{R2,t} * V1_{R2,t})$$

I tabell 3 och 4 nedan redovisas simuleringarna. Resultatet avser medelvärdet av de skattade parametrarna. Det vill säga, varje indextal har kunna beräknats 1000 gånger. Medelvärdet av dessa finns för 48 månader och medelvärdet av dessa medelvärden redovisas i tabellen nedan.

Därtill redovisas medelvärdet av standardavvikelseerna och maximum och minimum. I kolumn 1 och 2 redovisas först medelvärden etc. avseende de två regionala indexserierna som bygger upp den nationella serien. Nästkommande kolumner visar först de fyra beståndsviktade indexserierna och därefter de fyra transaktionsbaserade.

Tabell 3. Simuleringsresultat 1-8 (medelvärde per månadsskattning och dess standardavvikelse).

	R1	R2	V1	V2	V3	V4	V5	V6	V7	V8
Simulering 1										
Medel	70,40	129,45	99,92	107,91	111,50	109,56	99,92	107,92	111,45	109,57
Max	86,29	152,90	115,02	125,61	130,93	128,06	115,02	125,71	132,08	128,30
Min	55,52	108,47	87,09	92,60	94,73	93,58	87,09	92,38	94,10	93,37
St.dev	4,97	6,87	4,26	5,11	5,54	5,31	4,26	5,15	5,88	5,34
Simulering 2										
Medel	70,51	130,17	100,34	109,07	113,06	110,89	100,34	109,06	112,98	110,88
Max	126,11	201,70	139,05	163,90	173,98	168,40	139,05	163,56	174,38	168,16
Min	41,65	85,71	71,46	75,81	77,89	76,85	71,46	75,94	77,70	76,94
St.dev	12,25	17,62	10,80	13,20	14,34	13,73	10,80	13,20	14,46	13,73
Simulering 3										
Medel	70,66	129,90	100,28	108,27	111,85	109,92	100,28	108,25	112,17	109,89
Max	88,60	154,16	117,14	126,78	131,50	128,94	117,14	127,82	137,69	130,45
Min	57,20	110,52	88,39	94,20	95,53	95,26	88,39	93,61	93,02	94,24
St.dev	4,87	6,93	4,21	5,12	5,58	5,33	4,21	5,31	7,24	5,49
Simulering 4										
Medel	71,14	129,58	100,36	108,73	112,15	110,47	100,36	108,80	112,51	110,53
Max	121,57	204,55	141,22	164,85	173,57	169,78	141,22	165,14	179,08	170,21
Min	36,86	88,86	71,05	76,20	79,07	77,58	71,05	76,29	77,59	77,69
St.dev	12,70	16,81	10,83	12,68	13,49	13,14	10,83	12,78	14,50	13,22
Simulering 5										
Medel	70,46	129,60	100,03	107,72	111,22	109,32	100,03	107,72	111,17	109,33
Max	86,45	155,20	114,81	127,44	131,82	129,46	114,81	127,42	132,76	129,52
Min	57,30	109,22	86,36	92,08	94,74	93,36	86,36	91,77	93,73	93,19
St.dev	4,89	7,01	4,33	5,21	5,65	5,41	4,33	5,25	6,00	5,44
Simulering 6										
Medel	70,18	129,72	99,95	108,37	112,28	110,15	99,95	108,36	112,23	110,15
Max	15,76	199,24	141,37	163,14	172,05	167,18	141,37	163,16	173,59	167,34
Min	41,15	85,89	71,13	76,91	78,53	77,68	71,13	76,73	78,39	77,74
St.dev	11,99	17,60	10,76	13,12	14,25	13,64	10,76	13,12	14,38	13,64
Simulering 7										
Medel	70,45	129,24	99,84	107,46	110,93	109,05	99,84	107,53	111,31	109,11
Max	89,22	153,05	114,61	125,37	130,54	127,66	114,61	126,45	136,78	128,61
Min	55,12	110,29	87,09	93,34	95,66	94,35	87,09	92,85	92,54	94,27
St.dev	4,89	6,88	4,24	5,09	5,53	5,29	4,24	5,27	7,25	5,44
Simulering 8										
Medel	70,18	129,82	100,00	108,52	112,36	110,27	100,00	108,57	112,71	110,31
Max	121,24	203,05	141,02	164,84	173,09	168,97	141,02	166,78	178,18	170,41
Min	39,48	83,66	73,24	77,48	79,29	78,37	73,24	77,21	78,32	78,30
St.dev	12,70	17,48	10,72	12,98	14,08	13,48	10,72	13,13	14,89	13,59

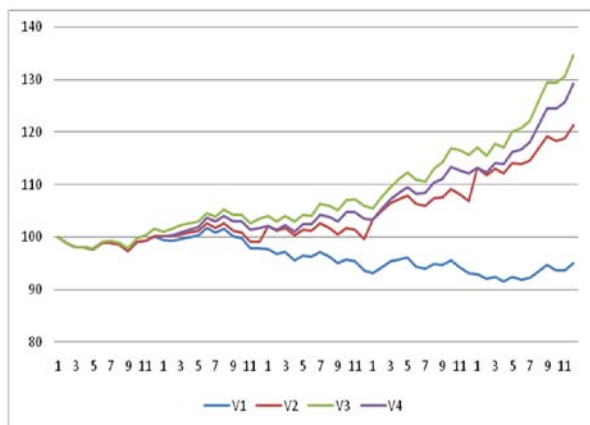
Tabell 4. Simuleringsresultat 9-16 (medelvärde per månadsskattning och dess standardavvikelse).

	R1	R2	V1	V2	V3	V4	V5	V6	V7	V8
Simulering 9										
Medel	70,62	188,75	129,68	153,57	161,77	157,46	129,68	153,56	161,63	157,44
Max	87,25	218,04	145,85	177,92	188,91	183,11	145,85	178,49	191,36	183,38
Min	55,18	163,27	113,38	132,45	138,84	135,53	113,38	132,53	136,25	135,24
St.dev	4,85	9,06	5,16	7,26	8,04	7,63	5,16	7,34	8,69	7,68
Simulering 10										
Medel	71,02	187,98	129,50	153,55	161,89	157,49	129,50	153,54	161,76	157,49
Max	123,86	280,40	176,94	232,55	248,44	240,14	176,94	233,55	249,31	240,94
Min	41,38	127,49	95,76	105,14	109,88	107,34	95,76	105,39	109,55	107,52
St.dev	12,46	22,52	12,74	18,14	20,08	19,07	12,74	18,17	20,32	19,08
Simulering 11										
Medel	70,33	188,11	129,22	153,06	161,24	156,94	129,22	153,16	161,71	157,03
Max	87,42	221,16	147,54	179,61	190,30	184,70	147,54	179,45	198,99	184,85
Min	56,36	162,18	113,47	132,49	138,33	135,41	113,47	130,76	131,46	133,57
St.dev	4,97	9,04	5,23	7,25	8,02	7,61	5,23	7,65	11,09	7,95
Simulering 12										
Medel	71,31	187,85	129,58	153,45	160,31	157,36	129,58	153,56	160,72	157,44
Max	118,44	268,59	174,52	216,50	231,60	224,68	174,52	220,73	239,16	227,86
Min	38,42	132,68	96,78	109,49	113,21	111,76	96,78	108,93	109,72	111,29
St.dev	12,28	22,17	12,60	17,83	19,38	18,74	12,60	18,06	20,75	18,90
Simulering 13										
Medel	70,39	188,20	129,30	152,63	160,70	156,45	129,30	152,64	160,59	156,46
Max	87,80	222,62	147,99	180,52	191,71	185,94	147,99	181,15	193,15	185,59
Min	57,77	160,68	112,51	130,84	136,56	133,62	112,51	130,07	134,72	133,31
St.dev	4,90	9,34	5,32	7,49	8,28	7,87	5,32	7,57	8,94	7,92
Simulering 14										
Medel	70,85	186,84	128,59	152,10	160,25	155,94	128,84	152,12	160,16	155,97
Max	120,08	271,48	174,59	220,71	236,49	227,99	174,59	221,53	237,62	228,64
Min	41,04	127,15	95,07	105,16	108,54	106,71	95,07	105,34	107,99	106,79
St.dev	12,50	22,74	13,09	18,20	20,10	19,10	13,09	18,24	20,36	19,13
Simulering 15										
Medel	70,36	188,36	129,36	152,76	160,86	156,59	129,36	152,79	161,27	156,62
Max	87,05	219,82	145,83	178,60	189,92	184,09	145,83	179,96	197,42	185,31
Min	57,45	161,81	114,83	132,61	137,68	135,09	114,83	130,21	131,06	133,72
St.dev	4,82	9,06	5,11	7,21	8,01	7,59	5,11	7,60	11,12	7,94
Simulering 16										
Medel	70,24	189,30	129,77	154,12	162,53	158,09	129,77	154,11	162,85	158,06
Max	124,20	297,76	185,11	245,94	262,15	254,22	185,11	248,88	265,34	254,41
Min	41,15	129,38	94,76	108,47	111,93	110,06	94,76	108,01	109,04	109,44
St.dev	12,76	22,64	13,23	18,24	20,10	19,11	13,23	18,46	21,51	19,28

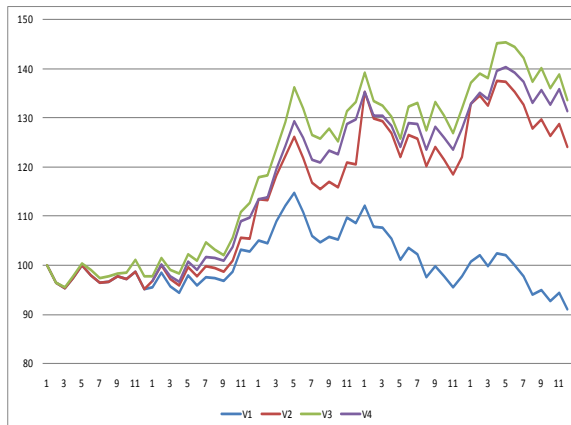
## 5. Tolkning av simuleringsresultat

I nedanstående figurer visas några realiserade indexserier, det vill säga några alla de 16000 indexserier som har skattats. I figurerna visas endast de beståndsbaserade vikterna. Det är tydligt att simulering 1 och simulering 9 har genererat betydligt jämnare serier där volatiliteten är lägre. Simulering 8 och 16 visar på serier med hög volatilitet.

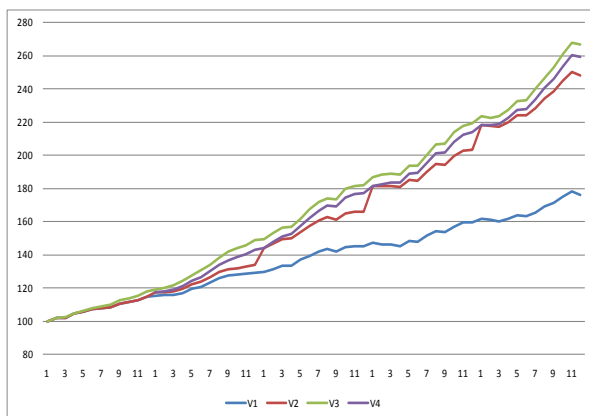
Figur 3. Simulering 1.



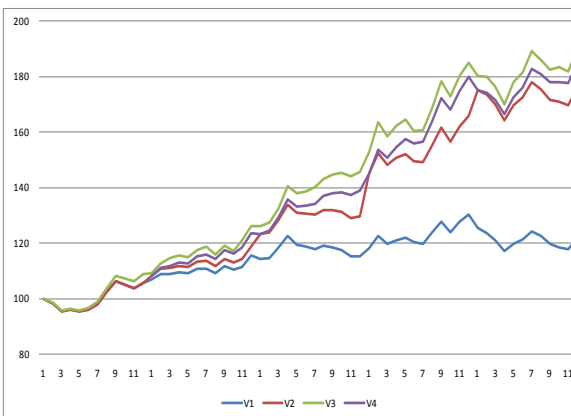
Figur 4. Simulering 8.



Figur 5. Simulering 9.



Figur 6. Simulering 16.



Om vi återgår till tabell 3 och 4 kan man konstatera att överlag är skillnaderna små mellan beståndsbaserade och transaktionsbaserade vikter. Den stora skillnaden består i om man uppdaterar vikterna eller ej. Viktningsprocedurerna 1 och 5 (som innebär att ingen uppdatering sker överhuvudtaget) skiljer sig avsevärt från de som uppdateras. Störst är skillnaden mellan dessa och de som uppdateras varje månad. Orsaken är naturligtvis att en växande marknad med högre priser egentligen borde väga tyngre i det komposita indexserien då de står för en högre andel av det totala värdet av bostadsbeståndet. Standardavvikelsen är generellt lägre i de index där viktningen inte uppdateras.

### *Prisantaganden*

I de simuleringar där vi antar en snabbare prisökningstrend i en av regionerna (simulering 9-16) kan man se att skillnaden mellan beståndsbaserat och transaktionsbaserat inte är större jämfört med att vi antar en långsammare ökning. Den stora skillnaden är återigen mellan de index där viktningen uppdateras och där de inte uppdateras. Den skillnad som man eventuellt

kan observera är att standardavvikelsen avseende indextalen där viktningarna uppdateras varje månad ökar något i den transaktionsbaserade viktningen.

En ökning i pris volatiliteten ökar av naturliga skäl standardavvikelsen i skattade indextalen. Däremot påverkas inte slutsatsen angående valet mellan transaktions- och beståndsbaserade viktningar. Skillnaden mellan de olika viktningmetoderna blir heller inte större om vi antar en snabbare tillväxt i form av prisökning i en av regionerna.

### ***Beståndsantaganden***

Beståndsantagandet har relativt liten påverkan på indextalen. Isolerat skulle man nog kunna konstatera att det inte påverkar valet mellan viktningmetod.

### ***Omsättningsantaganden***

Omsättningsantagandet är det som av naturliga skäl har störst påverkan på skillnaden mellan bestånds- och transaktionsbaserad viktning. Hög slumpterm avseende antalet transaktioner påverkar föga överraskande det transaktionsbaserade månadsindexet mest. Framförallt i kombination med en kraftig tillväxt i en av regionerna ser ut att öka volatiliteten i skattningen relativt mycket. Däremot ser det ut som om en stor volatilitet i prisutvecklingen motverkar omsättningseffekten. Det vill säga en ökad slump avseende antalet transaktioner och en regionsförstoring ser inte ut att påverka de transaktionsbaserade månadsindexserierna lika mycket om prisvolatiliteten är hög jämfört med om den är låg.

### ***Slutsats om viktning***

Vi har antagit relativt stora skillnader mellan regionernas prisutveckling, beståndsutveckling och omsättningshastighet. Trots det är det relativt små skillnader om man använder beståndet eller antalet transaktioner i viktningen. På basis av simuleringarna kan man dra slutsatsen att en transaktionsbaserad viktning bör kunna vara en god metod för viktning. Man bör vara försiktig om variationen i omsättningen varierar kraftigt från en månad till en annan. Det är dock inte samma sak som att säga att vi har en stor säsongsvariation när det gäller antalet transaktioner då det drabbar alla regioner samtidigt. Här avses om det är en stor variation i omsättningen i en region men den är oberoende av omsättningen i de andra regionerna. Det skulle tala för att använda sig av ett index som baseras på en genomsnittlig omsättningshastighet mätt i antalet transaktioner. Simuleringarna kan dock inte visa om ett 12-månaders rullande medelvärde är att föredra framför exempelvis ett 6-månaders medel. Simuleringarna visar dock att ett rullande medelvärde effektivt minskar standardavvikelsen av indexskattningarna.

## **6. Antalet regioner ingående i de komposita indexen**

I detta avsnitt kommer vi att analysera vilka regioner, och därmed antalet regioner, som skall ingå i det nationella indexet. Utgångspunkten är att hela Sverige skall vara representerat. Det innebär att vi inte beräknar ett index av typen S&P/Case-Shiller som endast består av 20 städer i USA. Deras kompositindex är mer likt S&P500 eller OMXS30 medan vårt index snarare kommer att vara mer likt OMX Stockholm PI. Nu är det dock inte möjligt att exempelvis inkludera samtliga Sveriges kommuner då det inte är möjligt att skatta index i samtliga kommuner i Sverige. För det första så finns det inte bostadsrätter i samtliga kommuner, för det andra så

kanske det inte förekommer några transaktioner vissa månader eller så förekommer det transaktioner men det är väldigt få. Därmed står man inför valet att välja vilka som skall ingå.

Det vi inte har övervägt är att skatta en hedonisk prisekvation för hela Sverige och därmed skatta ett nationellt prisindex. Anledningen till detta är det är högst osannolikt att alla implicita priser avseende de värdepåverkande attributen är konstanta mellan olika regioner.

### **Beslutskriterier avseende val av periodisering av index.**

I vissa regioner kommer således datamaterialet att vara tunt, det vill säga få antal bostadsrätter omsätts på marknaden. Det kommer att innebära att enskilda skattningar av indexvärden inte kommer att vara statistiskt signifikanta. Det kan då finnas skäl att istället för att skatta ett index per varje månad skatta index per varje kvartal/tertiäl eller till och med endast en gång om året.

Valet av periodisering måste baseras på en objektiv bedömning vad som är en relevant periodisering. Vårt beslutskriterium skulle kunna baseras på förändring av modellens förklaringsgrad. Det innebär att förklaringsgraden med och utan tidsdummies (tidsdummies) kommer att jämföras. Om förändringen i förklaringsgrad är signifikant så är det en indikation att vald periodisering är relevant. Om flera olika periodiseringar är signifikanta så väljs den modell med högst adjusted R-square (justerat  $R^2$ ). Valet av periodisering kommer således att ske i följande steg:

1. Skatta hedonisk prisekvation med tidsdummies (månad/kvartal/tertiäl/år). Spara R-square.
2. Skatta hedonisk prisekvation utan tidsdummies. Spara  $R_*^2$ .
3. Beräkna F-värdet (se ekvation nedan). Om F-värdet är större än det kritiska värdet är förändringen i förklaringsgrad ( $R^2$ ) signifikant.

$$F[J, n - k - 1] = \frac{(R^2 - R_*^2)/J}{(1 - R^2)/(n - k - 1)}$$

Där  $J$ =antal tidsdummies,  $n$ =antal observationer och  $k$ =antal oberoende variabler samt  $R^2$ =förklaringsgrad i modellen med tidsdummies och  $R_*^2$ =förklaringsgrad i modellen utan tidsdummies.

4. Alternativt skulle den modell som uppvisar den högsta förklaringsgraden kunna väljas, det vill säga om både en kvartalsmodell och en månadsmodell är signifikanta, väljs den modell med högst förklaringsgrad (justerat  $R^2$ ).

Likson "adjusted R-square" (justerat  $R^2$ ) tar AIC (Akaike Information Criterion) och BIC (Bayesian Information Criterion) hänsyn till den "trade-off" som är mellan antal oberoende variabler i modellen och antalet observationer, det vill säga antalet frihetsgrader. När det gäller både AIC och BIC så tar dessa mer hänsyn till den minskade antalet frihetsgrader som äger rum när fler oberoende variabler förs på modellen än vad justerat  $R^2$  gör. BIC är det mått som tar störst hänsyn till detta även om både AIC och justerat  $R^2$  gör det. Inget av måtten är självklart att föredra framför det andra. Båda har sina för- och nackdelar (Diebold, 2003). Vissa undersökningar tyder dock på att BIC är det som är att föredra (Kabaila, 2002) även om AIC ser ut att vara bättre än BIC i stora dataset (Shibata, 1981). Justerat  $R^2$ , AIC och BIC kommer att analyseras.

Samtliga mått ovan är så kallade "in-sample" baserade. Det är inte nödvändigtvis så att dessa mått är de som är mest lämpade. Dels finns det vissa mått som används vid "non-nested model selection" som kan användas, dels kan man basera sitt val på "out-of-sample" mått. Med det sistnämnda väljer man den specifikation som är "bäst" när det gäller att prognostisera värden på enskilda objekt.

Vi har valt att analysera två stycken vanligt använda metoder när det gäller att ta fram den "bästa" specifikationen. Den ena bygger på Davidsson och Mackinnon (1981), det så kallade J-testet, och den andra på Davidsson och xxx (xxxx). Utan att gå in på detalj så jämförs två olika specifikationer med olika uppsättning av oberoende variabler i J-testet alternativt Cox-Pesaran testet (se Cox, 1961 och 1962 samt Pesaran 1974).

I det andra testet så skapar man en "nested model" en så kallad supermodell som inkluderar samtliga oberoende variabler och man utför sedan ett F-test enligt ovan (Greene, 2000). Vid sidan om dessa kommer även modellernas prognosegenskaper att testas genom att "out-of-sample" prognoser tas fram. Det åstadkommes genom att vi drar ett urval från datamaterialet som nyttjas för att skatta modellerna och som därefter används för att skatta priserna på de objekt som inte ingår i urvalet. Därefter jämförs prognosen med faktiskt försäljningspris. Den specifikation som minimerar prognosfelet är den modell som väljs. Måttet som har använts är RSME (*Root Square Mean Error*).

Innan vi analyserar samtliga kommuner i Sverige med avseende på periodisering kommer två exempel att visas. Det ena avser Uppsala och där kommer vi att jämföra de olika "in-sample" metoderna, det vill säga justerat  $R^2$  samt AIC och BIC. Därefter kommer vi att presentera exemplet Eskilstuna där samtliga metoder kommer att presenteras.

### **Exempel 1 - (Uppsala kommun församling 1)**

En hedonisk modell skattas där endast bostadsyta, antal rum, ålder och månadsavgift ingår som förklarande variabler. Därefter skattar vi en modell där 48 stycken månadsdummies inkluderas och slutligen en modell där 16 stycken kvartalsdummies inkluderas istället för månadsdummies. I nedanstående tabell redovisas förklaringsgraden i respektive modell.

Tabell 5. Regressionsresultat alla objekt

	<b>Ingen tid</b>	<b>Månad</b>	<b>Kvartal</b>
<b>R<sup>2</sup></b>	0,6849	0,7525	0,7487
<b>R<sup>2</sup>-adj</b>	0,6845	<b>0,7490</b>	0,7474
<b>AIC</b>	-176,57	-921,27	<b>-985,28</b>
<b>BIC</b>	-145,48	-597,85	<b>-854,66</b>
<b>Antal observationer</b>	3713	3713	3713
<b>Antal oberoende variabler</b>	4	52	20
<b>F-värde</b>	-	20,82	59,48
<b>Kritiskt värde</b>	-	1,36	1,64

Modellen som inte tar hänsyn till tid förklarar 68,49 procent av variationen i pris med endast 4 stycken oberoende variabler. Modellen där månadsdummies har inkluderats förklarar 75,25 procent av variationen, det vill säga en ökning med 6,76 procentenheter. Ökningen i förklaringsgrad har kommit tillstånd då 48 oberoende variabler har inkluderats i modellen.

Genom att endast lägga på 16 kvartalsdummies har modellens förklaringsgrad ökat med 6,38 procentenheter. Är då dessa förändringar i förklaringsgrad statistiskt signifikanta? Det testar vi med att beräkna F-värdet. F-värdet avseende månadsindex är lika med:

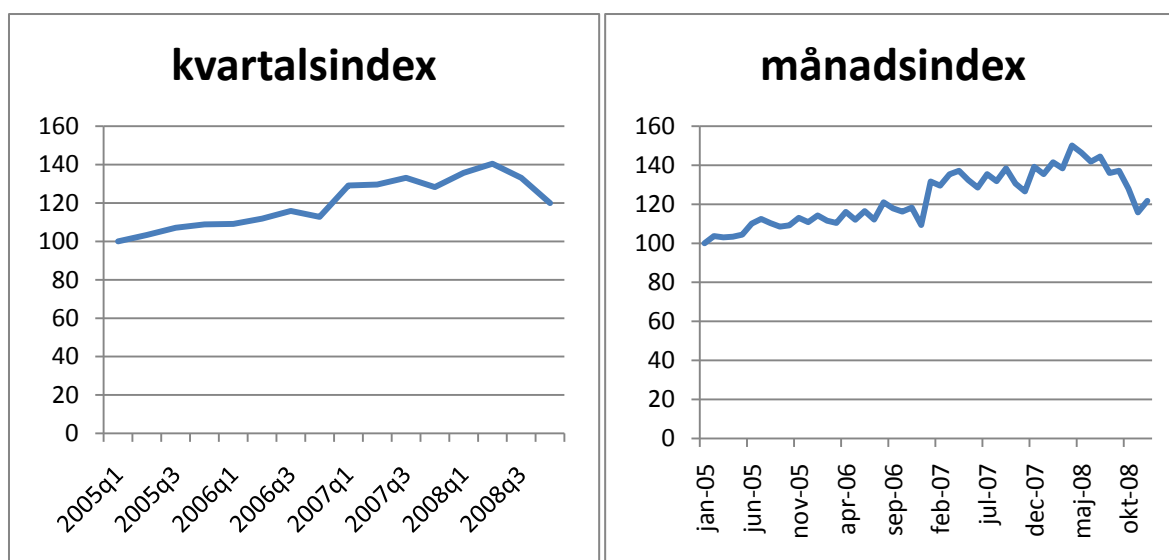
$$F[48,3713 - 52 - 1] = \frac{(0,7525 - 0,6849)/48}{(1 - 0,7525)/(3713 - 52 - 1)} = 20,82$$

Det kritiska värdet är lika med 1,45, vilket här innebär att vi kan förkasta hypotesen att inkluderandet av månadsdummies inte har någon förklaringskraft.

Motsvarande beräkning av *F*-värdet för modellen med kvartalsdummies är 59,48, vilket innebär att inte heller modellen avseende kvartalsdummies kan förkastas (kritiskt värde är här lika med 2,07). Man kan notera att förändringen i förklaringskraft i kvartalsmodellen är signifikant på en högre nivå.

Eftersom båda modellerna är signifikanta väljs månadsmodellen om kriteriet hade varit att ta den modell som har högst justerat *R*<sup>2</sup>. I detta fall månadsmodellen. Om däremot AIC eller BIC används som kriterium kommer kvartalsmodellen att användas. Kvartalsindex och månadsindex avseende Uppsala kommun (den största innerstadsförsamlingen) redovisas i nedanstående två figurer.

Figur 7. Kvartals och månadsindex i Uppsala (alla objekt)



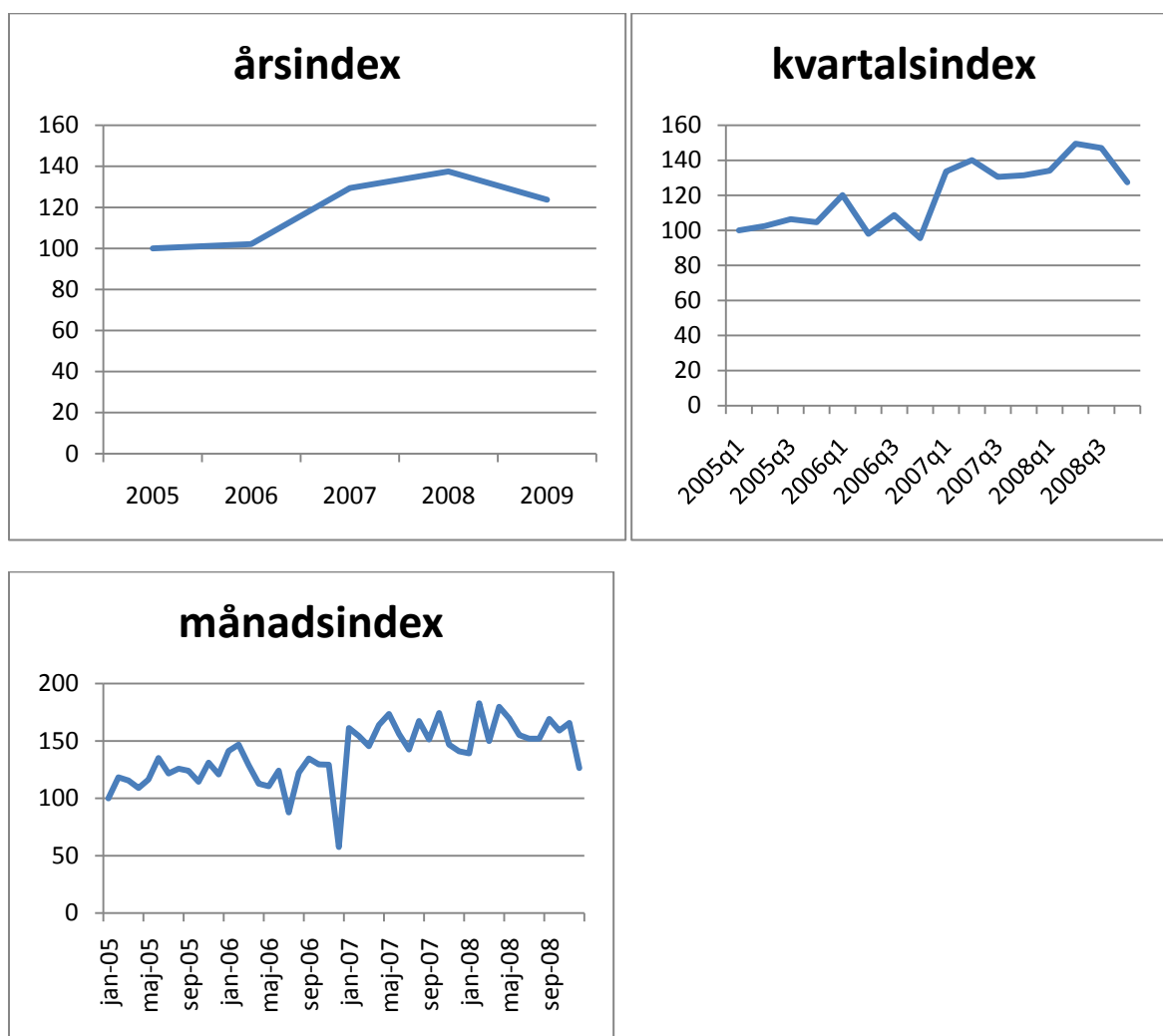
Motsvarande beräkningar har gjorts för bostadsrätter med exakt 4 rum och kök. Resultatet presenteras i nedanstående tabell.

Tabell 6. Regressionsresultat fyra rum och kök.

	Ingen tid	Månad	Kvartal	År
<b>R2</b>	0,3037	0,5511	0,4716	0,4477
<b>R2-adj</b>	0,2962	<b>0,4579</b>	0,4338	0,4336
<b>AIC</b>	123,04	89,35	76,12	<b>64,83</b>
<b>BIC</b>	137,66	275,81	149,24	<b>94,08</b>
<b>Antal observationer</b>	286	286	286	286
<b>Antal oberoende variabler</b>	3	49	19	7
<b>F-värde</b>	-	2,83	6,21	22,29
<b>Kritiskt värde</b>		1,42	1,68	1,62

Även i detta fall faller valet på det hedoniska indexet med månadsperiodisering om justerad R<sup>2</sup> används som mått. Om däremot AIC eller BIC används kommer kvartalsindexet att föredras, det vill säga eftersom kvartalsmodellen redovisar de lägsta AIC och BIC-värdena. I detta fall är det tom med så att en årsmodell är att föredra framför en kvartalsmodell. I nedanstående figurer redovisas årsindex, kvartalsindex och månadsindex avseende fyrrummare i Uppsala kommun, församling 1. Enligt BIC kriteriet är BIC att föredra.

Figur 8. Års- och kvartalsindex i Uppsala kommun (fyra rum och kök).



## Exempel 2 – Eskilstuna kommun

Eskilstuna kommun är en kommun där det har skett 1444 transaktioner, vilket motsvarar endast 25 försäljningar i genomsnitt per månad. Frågan är nu om det är möjligt att skatta ett månadsindex för Eskilstuna kommun. Rent tekniskt är det möjligt då vi har observationer för varje månad, men det troliga är att standardavvikelsen kommer att bli hög. Rent statistiskt kommer vi att testa sex olika metoder.

1. Justerat  $R^2$
2. AIC/BIC
3. J-test
4. Cox-Pesaran test
5. Supermodell
6. RSME

Tabell 7. Jämförelse av resultat.

	Justerat $R^2$	AIC	BIC	J-test	Cox-Pesaran test	Supermodell	RSME
<b>Månadsmodell</b>	0,7365	722,78	1144,80	0,00	0,47	-	198927
<b>Kvartalsmodell</b>	0,7329	707,23	934,06	0,00	0,00	-	206328
<b>Skillnad</b>	-	-	-			0,024	

Vad säger då detta: Justerat  $R^2$  visar på att högst förklaringsgrad får vi i den modell där vi skattar ett månadsindex. Däremot visar både AIC och BIC att kvartalsmodellen är bättre. Tyvärr visar  $J$ -testet på att det inte är möjligt att särskilja på modellerna (i tabellen återges  $p$ -värden). Däremot visar Cox-Pesaran på att månadsmodellen är att föredra (även är  $p$ -värden). Även supermodellen visar på att vi kan förkasta hypotesen att samtliga månadsskattningar är lika med noll, det vill säga en månadsmodell är att föredra. Hypotesen har förkastats med hjälp av ett  $F$ -test och  $p$ -värdet är det som visas i tabellen. Det sista måttet är en "out-of-sample" skattningen. Här har vi dragit ett slumpmässigt 80 procentigt urval där modellens skattningar har applicerats på de 20 procent som inte ingick i estimeringen. I detta fall är det 245 observationer som vi har skattat en prognos för dessa och jämfört dessa med det faktiska priset. Det genomsnittliga felet visas i tabellen ovan. Vi kan observera att det genomsnittliga felet är lägre i månadsmodellen än i kvartalsmodellen vilket skulle tas för intäkt att denna modell är bättre som prognosmodell.

Vilket mått skall man då välja? Naturligtvis finns det för och nackdel med samtliga mått. Vissa mått är enklare att förstå medan andra är mer exakta. Det som vi har tagit fasta på är modellens prognosegenskaper och då framför allt dess "out-of-sample" egenskaper. Därför skulle man kanske lägga lite större tilltro till RSME-måttet. Å andra sidan är resultatet från supermodellen intressant då den tittar på förändringen i förklaringsgraden och om den är statistiskt signifikant eller ej.

Det vi emellertid kommer att göra är att skatta samtliga mått per kommun (dock ej i detta skede  $J$ -test och Cox-Pesaran test). I de fall där testen överensstämmer är naturligtvis problemet mindre, men i de fall där testen visar på olika periodiseringar är det naturligtvis svårare att dra

några slutsatser. I nedanstående avsnitt kommer tillvägångsättet vid skattningen av ett Sverigeindex avseende bostadsrätter att presenteras.

### ***Tillvägångsätt vid skattning av Sverigeindexet***

Att enbart beräkna ett Sverigeindex med de regioner där månadsindex är att föredra rent statistiskt innebär att inte hela Sverige ingår vid beräkningen av Sverigeindexet vilket är olyckligt. Naturligtvis kan vi beräkna ett HOX5, men då är det viktigt att det inte kallas för något annat än HOX5. Nu kommer det dock inte att vara stora skillnader mellan ett Sverigeindex och exempelvis HOX5, så valet har liten praktisk betydelse.

Om hela Sverige skall ingå skall dock även de kommuner där inte månadsindex fungera ingå. Vi har här valt att aggregera ihop dessa kommuner länsvis. Det innebär att de kommuner i Stockholms län där inte ett månadsindex kan beräknas slås ihop till ett Stockholms länsindex exklusive de kommuner som faktiskt har ett eget månadsindex. Fungerar inte heller det, så slås län ihop i syfte att få sammanhängande regioner med ett månadsindex. Fördelen är att alla regioner (kommuner) ingår i ett länsindex eller ett sammanslaget länsindex. Nackdelen är dock att den egentliga prisutvecklingen i en region inte följer den prisutveckling som länet.

Målet har varit att samtliga regioner där ett månadsindex är att föredra framför ett kvartalsindex skall ingå som en separat region i beräkningen av det nationella indexet. Således har vi börjat undersöka varje kommun för sig avseende om månadsindex eller kvartalsindex är att föredra. Resultatet redovisas i tabellen nedan. Dessa beräkningar har gjorts för de 20 största kommunerna mätt i antal bostadsrättstransaktioner. En utförligare tabell återges i appendix där justerat  $R^2$  presenteras tillsammans med AIC och BIC. Det som kan konstateras är att väldigt få kommuner "kvalar in" när man använder sig av AIC eller BIC. Däremot använder man sig justerat  $R^2$  så ingår nästan samtliga kommuner. Använder man istället de andra metoderna såsom supermodellen och "out-of-sample" beräkningar får man ett delvis annat resultat. Även det åskådliggörs i tabellen nedan.

Tabell 8. Periodisering i de 20 största kommunerna i Sverige.

	Nr. obs.	R <sup>2</sup>		AIC		"Supermodell"		RSME	
		M	K	M	K	M	K	M	K
Stockholm	36842	✓		✓		✓		✓	
Göteborg	19483	✓		✓		✓		✓	
Uppsala	10502	✓				✓		✓	
Malmö	7024	✓				✓		✓	
Solna	3820	✓		✓		✓		✓	
Nacka	3474	✓		✓		✓		✓	
Linköping	3381	✓				✓		✓	✓
Järfälla	3218	✓				✓		✓	
Västerås	3427	✓						✓	
Örebro	2348		✓					✓	
Lund	2803	✓				✓		✓	
Haninge	2592	✓				✓		✓	
Umeå	2524	✓						✓	
Gävle	2481	✓		✓		✓		✓	
Helsingborg	2403	✓						✓	
Jönköping	2371	✓						✓	
Karlstad	2190	✓						✓	
Täby	1808	✓				✓		✓	
Mölnadal	1730	✓						✓	
Huddinge	1696	✓		✓		✓		✓	
Sundsvall	1676	✓						✓	
Norrköping	1636	✓				✓		✓	
Luleå	1630	✓						✓	
Östersund	1620		✓					✓	
Nyköping	1609	✓				✓		✓	
Borås	1578	✓						✓	
Sollentuna	1558	✓						✓	
Botkyrka	1545		✓					✓	
Sundbyberg	1456	✓				✓		✓	
Eskilstuna	1444	✓				✓		✓	
Kalmar	1318	✓						✓	
Norrtälje	1299	✓						✓	
Sigtuna	1289		✓					✓	
Halmstad	1232	✓				✓		✓	
Södertälje	1177	✓						✓	
Lidingö	1145		✓					✓	
Upplands-Bro	1097	✓				✓		✓	
Tyresö	1071	✓						✓	
Danderyd	776	✓				✓		✓	
Falkenberg	430	✓						✓	
<b>Antal kommuner</b>		<b>35</b>		<b>6</b>		<b>20</b>		<b>39</b>	

Resultatet spretar lite. Tittar man enbart på justerat  $R^2$  så ingår nästan samtliga undersökta kommuner. Om man däremot använder sig av AIC så ingår nästan inga överhuvudtaget. Det som ser ut att vara en medelväg är att gå på den så kallade "Supermodellen". Testet vi där gör är att se om förändringen i  $R^2$  är statistiskt signifikant på en 5 procentig nivå ( $F$ -test). I 20 kommuner är månadsindex att föredra framför ett kvartalsindex. Det kan jämföras med 35 kommuner om man går på absolutvärdet på justera  $R^2$ .

Det som inte verkade fungera som tänkt är "out-of-sample" metodiken. Den indikerar nästan uteslutande att månadsmodell är att föredra oavsett om dessa är signifikanta eller ej. Med tanke på dess avvikande resultat har vi valt att inte gå vidare med den metodiken.

## **7. Slutsatser och rekommendationer**

För att beräkna ett nationell kompositindex krävs att man dels fastställer hur viktningen av de enskilda regionala indexserierna skall gå till, och dels fastställer hur många regioner som skall ingå i det nationella indexet.

När det gäller valet av viktningssmetod så rekommenderar denna studie beståndsbaserad värdeviktning. Detta innebär att en regions vikt beror på prisnivån i regionen och antalet fastigheter/bostadsrätter i regionen i förhållande till det totala värdet av alla fastigheter/bostadsrätter i Sverige. Vikterna bör i princip uppdateras varje månad. Monte Carlo simuleringarna visar att det är relativt små skillnader om man använder beståndet eller antalet transaktioner i viktningen. På basis av simuleringarna kan man sluta sig till att en transaktionsbaserad viktning är rimlig att använda sig av. Man bör vara försiktig om variationen i omsättningen varierar kraftigt från en månad till en annan. Här avses om det är en stor variation i omsättningen i en region men den är oberoende av omsättningen i de andra regionerna. Det skulle tala för att använda sig av ett index som baseras på en genomsnittlig omsättningshastighet mätt i antalet transaktioner. Simuleringarna kan dock inte visa om ett 12-månaders rullande medelvärde är att föredra framför exempelvis ett 6-månaders medel. Simuleringarna visar dock att ett rullande medelvärde effektivt minskar standardavvikelsen av indexskattningarna.

Valet av antalet regioner som ska ingå i det komposita indexet är mer ett affärsmässigt val (med andra ord, valet av viktningssmetod är av större intresse ur ett statistiskt perspektiv). Dock kan det visa sig att vissa ingående regionala indexserier inte är lämpade att ur en statistisk synvinkel att redovisas. Valet av periodisering i respektive grupp kan man analysera utifrån den statistiska modellens prediktionsförmåga. Av de 40 största kommunerna (mätt i antal bostadsrättsförsäljningar) är det endast 20 som är lämpade att ha en periodisering på månad. Resterande kommuner är mer lämpade att ha en kvartalsperiodisering.

## Referenser

Gerlach, S. och Peng, W. (2005). Bank lending and property prices in Hong Kong. *Journal of Banking & Finance*, Vol.29, sid 461-481.

Levy, H. och Post, T. (2005). *Investments*. Prentice Hall, Financial Times.

Lum, S. K. (2004). Property price indices in the Commonwealth. *Journal of Property Investment & Finance*. Vol.22(1), 25-54.

Rappaport, J. (2007). Comparing Aggregate Housing measures. *Business Economics*, October 2007.

*S&P/Case-Shiller Home price Indices – Index Methodology*. S&P Indices, November, 2009.

Song, H-S och Wilhelmsson, M (2010). Improved price index for condominiums. Kommande i *Journal of Property Research*.

Villkor för aktieindex OMX30. I *Regelverk för OMX Derivatives Markets*, 4.8. Addendum OMX30.

Wilhelmsson, M. (2008). Housing Price Depreciation rates and Maintenance Level. *Journal of Housing Economics*. Vol. 17(1), 88-101.