



*Aina Holmøy, Randi Johannessen og
Leiv Solheim*

**Etablering av ny husleie-
statistikk (indeks) - en forstudie**

Notater

Innhold

1 Innledning	3
1.1 Bakgrunn og formål	3
1.2 Noen hovedresultater	4
1.3 Kort om kapitlene	4
2 Ny husleiestatistikk.....	5
2.1 Kjennetegn ved boforhold.....	5
2.2 Mulige kilder for husleiedata	5
2.2.1 Folke- og bolig tellingen 2001 (FOB2001)	5
2.2.2 Det sentrale folkeregisteret (DSF).....	6
2.2.3 Registeret over grunneiendommer, adresser og bygninger (GAB)	6
2.2.4 Kommunale boliger, studentboliger, profesjonelle utleiery.....	7
2.3 Screening	8
3 Utvalgsstørrelse.....	8
3.1 Nivå tall vs. endringstall	8
3.2 Variasjon og usikkerhet	8
3.3 Beregning av utvalgsstørrelsen.....	10
3.3.1 Notasjon.....	10
3.3.2 Nivå tall	11
3.3.3 Fra strata via gruppe til total tall.....	12
3.3.4 Indeks - endringstall	13
4 Oppsummering og anbefaling	15
4.1 Datakilde	15
4.2 Utvalg - størrelse og stratifisering.....	16
4.3 Datafangst og frafall	16
4.4 Avsluttende merknader	17
5 Litteraturliste	18
De sist utgitte publikasjonene i serien Notater	19

1 Innledning

1.1 Bakgrunn og formål

Statistisk sentralbyrå (SSB) gjennomfører en månedlig husleieundersøkelse. Denne brukes for det første til å beregne endringene i Konsumprisindeksen (KPI) som skyldes endringer i husleier i leiemarkedet. I Norge er den dominerende boform selveier. Det er særlige måleproblemer forbundet med å beregne prisendringer i dette markedet i KPI og dessuten beregne priser og verdiskapning i hussektoren for selveiere. I Norge har en valgt å benytte priser fra leiemarkedet til å beregne priser i eiemarkedet. Dette stiller store krav til metode og data da boligene i eiemarkedet og leiemarkedet er ganske ulike. Den nye loven om husleieavtaler (1999) fastslår at husleien ikke kan være høyere enn vanlig markedsleie og bare økes i samsvar med KPI. Først etter tre år er det mulig å justere leieprisen ytterligere og da i samsvar med gjengs leie. Den husleieundersøkelsen som SSB i dag her gir ikke et tilfredsstillende grunnlag for å estimere gjengs husleienivå eller utvikling. Gjengs leie er en ny rettslig standard og det forventes at det etter hvert vil komme presedensskapende avgjørelser i domstolene. For å støtte opp om denne prosessen er det ønsket av Regjeringen at det utvikles en ny husleieindeks som er mer omfattende og adekvat ved praktisering av husleielovens bestemmelser enn den husleiestatistikken som SSB produserer i dag. Det vil gjøre det lettere å fastsette hva som er gjengs leie i den enkelte sak i avgrensede geografiske områder.

Først og fremst er en slik indeks nødvendig for å få en forsvarlig og enhetlig praktisering av bestemmelsene om fastsetting og regulering av husleie til gjengs leie, men også andre samfunnsmessige hensyn vil nyte godt av den. En slik statistikk kan bli nyttig både ved utarbeidelsen av nasjonalregnskapet og ved internasjonale prissammenligninger. Den vil dessuten også kunne gi et bedre grunnlag for å vurdere bokostnadene for spesielle grupper.

Følgende oppdrag er gitt i Kommunal- og regionaldepartementets (KRD's) tildelingsbrev til Husbanken for 2003: " *Det er ønskelig å få etablert en husleieindeks for å bedre kunnskap om husleiemarkedet og utviklingen på dette delmarkedet*".

For å utrede muligheter og betingelser for en ny og bedre husleiestatistikk i Norge, satte Husbanken høsten 2003 ned en prosjektgruppe bestående av Åsmung Langsether fra NOVA og Per Melby fra Byggforsk. Utredningen skulle konsentrere seg om:

- klargjøring av formålet med en slik indeks
- kartlegging av mulige datakilder og vurdering av aktuelle datainnsamlingsmetoder
- gi en drøfting av kravene som må stilles til en ny husleieindeks og komme med anbefalinger på hvilken indekssløsning som bør velges
- gi en vurdering av hvor driftsansvaret for en slik indeks bør plasseres

Arbeidet ble påbegynt i november 2003 og resulterte i rapporten "Husleieindekser og husleiestatistikk" (NOVA Rapport 10/2004). I samråd med SSB anbefales det her at arbeidet med å etablere en ny husleiestatistikk inndeles i tre faser:

- Fase 1: Estimeringer på bakgrunn av foreliggende datasett/registre
- Fase 2: Screening for å etablere utvalg
- Fase 3: Oppstart av ny undersøkelse

SSB er i første omgang engasjert av Husbanken i gjennomføringen av forstudien som defineres som fase 1 innenfor det anbefalte opplegget. Denne fasen innebærer statistiske og metodemessige analyser av foreliggende datasett som inneholder leietagere. SSB skal i denne fasen:

- beregne variansen i dagens husleiestatistikk og på bakgrunn av dette estimere den statistiske usikkerheten for husleiemarkedet.
- gi mer presise anslag på hvilke strata det er hensiktsmessig å dele leiemarkedet inn i
- anbefale effektive utvalgsstørrelser for strataene
- avklare og dokumentere eventuell ny viten om det norske leiemarkedet med hensyn til hvilke faktorer som påvirker leiebeløpene, som estimeringene genererer

Dette notatet beskriver de analyser, resultater og anbefalinger som SSB har kommet fram til.

1.2 Noen hovedresultater

- Som datakilde for å etablere bruttoutvalget anbefales det å benytte en kombinasjon av Det sentrale personregisteret (DSF) og registeret over grunneiendommer, adresser og bygninger (GAB).
- Enkelte adresser i flerbolighus mangler fremdeles identifikasjon (leilighetsnummer). Særlig Oslo mangler mest på fullføring av leilighetsnummer. En økt kvalitet på leilighetsnummer vil medføre reduserte kostnader til datafangst.
- Sameie- og andelsboliger kan ikke i dag identifiseres fra registrene, men denne situasjonen vil bli bedre ved opprettelsen av et borettsregister
- Et bruttoutvalg på 15 000 er nødvendig for å oppnå et nettoutvalg på 5 000.
- Utvalget bør stratifiseres etter geografi, størrelse og hustype. Gitt et nettoutvalg på 5 000, blir tallene veldig usikre for små fylker. Et alternativ kan være å benytte landsdeler, eventuelt delt mellom tettbygd og ikke-tettbygd strøk, som geografisk nivå fremfor fylke.
- For å kunne produsere tall for de største byene og for en ytterligere inndelingen i områder innen disse, må det trekkes et tilleggsutvalg. Dersom det kreves en presisjon på ned mot 1 prosent innenfor områdene, må et slikt tilleggsutvalg være på omlag 2 500 boliger.
- Som størrelsesvariabel anbefales antall rom fordi antall rom gir en mer funksjonell inndeling enn boligareal målt i kvadratmeter.
- På grunnlag av de data vi har undersøkt kan det, gitt et utvalgsstørrelse på 5 000, publiseres nivåtall for landet med en usikkerhet på ned mot 1 prosent, mens nivåtall for de største fylkene kan gis med en sikkerhet innenfor 3-5 prosent.
- Nettoutvalget etableres gjennom en skjemabasert screening av bruttoutvalget.
- Datafangsten kan skje gjennom *computer assisted interview* (CAI).
- Utredning av faktorer som gir opphav til forskjeller i husleienivå og herunder utvikling av en hedonisk metode for prisindeksberegninger anbefales.
- Utredning av muligheten for å hente inn data fra profesjonelle utleiery (offentlige og private), i kombinasjon med datainnsamling blant de som leier av reint private utleiery.

1.3 Kort om kapitlene

I kapittel 2 vil vi gå nærmere inn på utfordringene rundt etablering av utvalg. Vi ser litt på mulige datakilder og fordeler/ulempene ved bruk av de ulike kildene. I kapittel 3 har vi gjort analyser på usikkerheten i data fra husleieundersøkelsen 2003, og vi ser nærmere på hvor stort utvalg en trenger for å lage en god husleieindeks. Vi viser at utvalgsstørrelsen i stor grad avhenger av hvilke krav man har til presisjon i resultatene. I kapittel 3 ser vi også litt på mulige inndelinger i strata. Det legges da vekt på inndeling i strata med varig innvirkning på husleienivået. I kapittel 4 gjør vi en oppsummering og kommer med anbefalinger.

2 Ny husleiestatistikk

Ifølge rapporten fra NOVA og Byggforsk er det i praksis bare Statistisk sentralbyrå (SSB) som kan stå for produksjon og publisering av en ny husleiestatistikk. En slik statistikk må, i mangel på adekvate registre, baseres på utvalgsundersøkelser. Det hersker imidlertid stor usikkerhet om hvor stort utvalg som kreves, hvilke variabler som bør legges til grunn for dekomponering av en slik statistikk, og hvilke datakilder som kan legges til grunn. Fra SSB's side er det derfor en forutsetning at det før en ny husleiestatistikk etableres, gjennomføres en forstudie for å avklare eksisterende usikkerhet/uklarhet så langt det er mulig.

2.1 Kjennetegn ved boforhold

Ifølge FOB 2001 er boligene fordelt på eier- og leietagere slik som tabell 1 nedenfor viser. Denne fordelingen bygger på hvor personer er registrert bosatt, noe som innebærer at mange av leieforholdene til studenter faller utenfor her siden studentene stort sett er registrert bosatt hos foreldrene. Det betyr at det i virkeligheten er flere leietagere enn hva tabell 1 viser.

Tabell 1. Boliger og bosatte i privathusholdningene fordelt på eier- og leieforhold. 3. november 2001. Prosent

Eier- og leieforhold	Boliger	Bosatte
Antallet totalt	1 961 548	4 485 945
Eies i alt	76,7	82,1
Selveier alene eller gjennom sameie	62,5	70,7
Gjennom borettslag eller aksjeselskap	14,1	11,5
Leies i alt	23,3	17,9
Av privatperson	13,0	10,1
Av boligselskap	2,5	1,6
Av kommunen	3,8	2,6
Som tjenestebolig	1,0	0,9
På andre vilkår	3,1	2,6

Med utgangspunkt i disse tallene er det 23,3 prosent av boligene som er målpopulasjon for husleiestatistikken. Disse utgjør 17,9 prosent av alle bosatte i private boliger ifølge FOB2001. Det betyr at dersom en trekker et tilfeldig utvalg av personer må en regne med at bare drøyt hver sjette av disse bor i en bolig som leies. Det blir derfor svært dyrt å trekke et stort tilfeldig utvalg for å finne fram til hvilke av disse som er leietagere.

2.2 Mulige kilder for husleiedata

Det finnes i dag ikke noe register over leietagere, som er målpopulasjonen i en husleiestatistikk. En står derfor overfor flere utfordringer når en skal etablere et utvalg over utleieboliger. Her skal vi se nærmere på noen mulige datakilder for etablering av et slikt utvalg.

2.2.1 Folke- og boligtellingsen 2001 (FOB2001)

For å etablere et utvalg over private utleieboliger, kan vi benytte oppdaterte filer fra Folke- og boligtellingsen (FOB) i 2001. Det er imidlertid et voksende problem med FOB-registeret. Det er ikke etablert tilfredsstillende rutiner for å holde populasjonen av leieboliger i registeret à jour, dvs. rutiner som tar vare på at noen boliger opphører å være utleieboliger og at det stadig vil være et tilslag til populasjonen av boliger. Nye utleieboliger vil ikke komme med i utvalget. Problemet kan delvis løses ved at det trekkes et tilleggsutvalg fra Det sentrale folkeregisteret (DSF), slik at vi får med noen nye utleieboliger på denne måten.

2.2.2 Det sentrale folkeregisteret (DSF)

Ifølge FOB2001 var under en fjerdedel av boligene utleieboliger, mens over tre fjerdedeler av boligene ble disponert gjennom et eierforhold. Siden husholdningene er gjennomgående mindre når boligen leies enn eies, kan vi forvente at mellom 15 og 20 prosent av et personutvalg bor i en bolig som leies. Regner vi et frafall på 20 prosent blant leietagerne, må man trekke et sted mellom 35 000 og 45 000 familiekjerner for å kunne etablere et utvalg på 5000 utleieboliger. Et tilfeldig personutvalg blant de over 16 år vil derfor være svært lite effektivt for å finne fram til et panel av leietagere som skal følges for å studere husleieutviklingen.

2.2.3 Registeret over grunneiendommer, adresser og bygninger (GAB)

Når en person eier en bolig alene skal dette i prinsippet være registrert i GAB gjennom at boligen er tinglyst. Ved å utnytte leilighetsnummer som identifiserer den som bor i en bolig (registrert bosatt), kan vi trekke ut informasjon om boligen eies av den eller de som bor i boligen. Dette er sannsynligvis ikke fullt ut dekkende for alle boliger siden en del selvstendig næringsdrivende kan eie boligen gjennom selskapet han eller hun driver sin næringsvirksomhet gjennom. Dette problemet kan trolig løses ved å koble person (fødselsnummer) mot foretak (organisasjonsnummer). For de som eier gjennom sameie er det vanskeligere å gjøre en kobling gjennom organisasjonsnummeret til sameiet og personen (personnummer) som er med i sameiet. Dette problemet vil vi gå videre med i en eventuell fase 2 for å undersøke om det er mulig å få til en løsning.

I dag finnes det ikke noe register over de som eier bolig gjennom borettslag eller aksjeselskap. Vi har derfor ingen mulighet for å identifisere eiere av borettslagsboliger. Det er imidlertid planer om å opprette et register over borettslag, som vil holde oversikt over denne gruppen som ifølge FOB2001 utgjør 14,1 prosent av boligene og 11,5 prosent av de bosatte i privathusholdningene.

Vi kan derfor konkludere med at det kun er i de tilfellene vi har en selveierbolig hvor identifikasjonen er eierens fødselsnummer på tinglysningen registrert i GAB, at vi kan fastslå at boligen er eid av minst en av de registrerte bosatte i denne boligen, altså ved å bruke leilighetsnummeret.

På grunnlag av data fra GAB på folketellingstidspunktet har vi koblet eierskap av boliger, dvs. med fødselsnummer som identifikasjon, med registrerte bosatte i boligene, fødselsnummer og leilighetsnummer. Vi finner at 49 prosent av de 1 961 548 boligene i tabell 1 ovenfor eies av en av de bosatte. Om differansen mellom de 62,5 prosent som ifølge FOB2001 er selveiere alene eller gjennom sameie og de 49 prosentene vi fant som eiere i GAB stemmer med de mulige forklaringer på avvik som vi beskrev ovenfor (sameie, næringsvirksomhet), har vi ikke kunnet undersøke.

Konklusjonen er imidlertid at dersom vi starter med å angi en målpopulasjon som er lik den delen av befolkningen som er bosatt i de 51 prosent av boligene der ingen av de bosatte er registrert som eier, slik kriteriet er brukt ovenfor, kan vi regne med at nærmere halvparten av disse er leietagere. I så fall bør et utvalg på omlag 15 000 kunne gi et godt grunnlag for å finne 5 000 leietagere. Anslaget baserer seg på at ca 10 000 svarer i screeningen og at to av tre som svarer er leietagere. Vi regner i tillegg et frafall på 20 prosent på selve kvartalsundersøkelsen.

To problemer må løses for at metoden skal fange opp alle leietagere i målpopulasjonen. For det første må alle boliger få tildelt et leilighetsnummer. Slik det er nå er ikke alle adresser i flerbolighus registrert i DSF/GAB. Manglende identifikasjon av adresse slår ut geografisk, da særlig Oslo mangler mest på fullføring av leilighetsnummer. Dette innebærer at vi får en større screening i Oslo enn i resten av landet, og dermed større kostnader.

Det andre problemet som må løses er avviket mellom formell bolig (registrert bosatt på) og faktisk bolig (der en har sin hovedsakelige døgnhvile). Avviket mellom faktisk og formell bosted gjelder særlig skoleelever og studenter og slår særlig sterkt ut i aldersgruppa 20-29 år. Mange studenter er registrert bosatt på foreldrenes bostedsadresse, og vil da bli ekskludert fra utvalget selv om de egentlig

bor i en leid bolig. For å illustrere problemet viser FOB2001 at av elev- og studentmassen på 104 000 personer som i 2001 søkte utdanningsstipend i Statens lånekasse og var registrert i utdanning i Norge høsten 2001, var 63 prosent registret bosatt utenfor skolekommunen mens de oppgav at de faktisk var bosatt i skolekommunen. En mulig løsning på dette problemet kan være å trekke et eget utvalg i aldersgruppa 20-29 år fra DSF og gjennomføre en screening blant disse for å få tak i skoleelever og studenter.

Dersom vi får løst problemene skissert over og ser framover i tid med et boretregister, er utsiktene gode for at vi med denne metoden kan trekke et utvalg som er tilnærmet lik målpopulasjonen, og som dermed kan brukes direkte i en husleieundersøkelse.

2.2.4 Kommunale boliger, studentboliger, profesjonelle utleiere

I prinsippet kan vi tenke oss husleiemarkedet inndelt i følgende utleiegrupper:

- Private utleiere
- Profesjonelle private utleiere
- Kommunale utleiere
- Studentsamskipnader

En tilnærming for å redusere utgiftene til datafangst kan være å beregne husleienivå og utvikling separat for disse utleiegruppene. Husleiedata for alle andre kategorier enn reint private utleiere hentes fra ulike kilder som studentsamskipnaden, fra kommuner, internett og fra utleiefirma. Dette innebærer at nettoutvalget kan begrenses til rene private utleiere. Gjennomsnittlig leienivå (eller indekser) for hver kategori vektet så sammen til et samlet husleienivå (eventuell indeks) for de strata som vi velger å dele markedet inn i.

Et problem med en slik tilnæringsmåte er at vi ikke har fullstendig oversikt over de ulike populasjonene og kan dermed få et problem ved vektning av husleiedata for disse ulike utleiekategoriene. Fra FOB2001 og NOVA-survey 2001 (jf. Langsether mfl. 2003) vet vi til en viss grad sammensetningen av utleiesektoren og størrelsen på ulike kategorier. Disse opplysningene kan benyttes til å vekte de ulike kategoriene. Problemet er at opplysninger om sammensetning langt fra er uttømmende. I tillegg må vi ha en løpende oppdatering av sammensetningen for å oppnå løpende vektning. Alternativt må vi anta at sammensetningen er noenlunde konstant og at vektene holdes på 2001-nivå. I tillegg vil det være overlapp mellom de ulike kategoriene; studenter leier ikke utelukkende gjennom studentskipnaden og sosial boligleie skjer ikke utelukkende gjennom boliger som kommunen selv eier.

Beregninger for 2001, basert på Levekårsundersøkelsen i 2001, viser stort samsvar i leiemarkedenes struktur i Oslo og i resten av landet. Beregninger basert på data fra folke- og bolig tellingen gir som resultat en større andel private profesjonelle utleiere i Oslo enn i resten av landet (21,2 % vs. 8,3 %). Begge datakildene gir estimater på offentlig utleieandel på anslagsvis 15-16 prosent både i Oslo og i landet for øvrig. Dette viser at det ved en slik tilnæringsmåte vil være behov for å hente inn data fra profesjonelle utleiere over hele landet.

Ifølge Langsether og Medby (2004) kan man tenke seg lokale løsninger på dette statistikkbehovet, men de vil kreve at det etableres permanente lokale organer som står for statistikkproduksjonen, og de bør - av hensyn til likebehandling av folk når det gjelder gjengs leie-problematikken - produsere en ev. slik statistikk etter standardiserte opplegg. En statistikk som produseres bare for noen lokale markeder, eller ev. bare for enkelte segmenter av leiemarkedet (f.eks. bare for noen utleiesektorer), introduserer en ulikhet mellom aktørene på ulike delmarkeder etter hvorvidt de har adgang til adekvat statistikk eller ikke, selv om den statistikken som er tilgjengelig på enkelte delmarkeder er laget etter samme lest. Langsether og Medby anser slike løsninger som både urealistiske og ressurskrevende i et nasjonalt perspektiv. I tillegg vil de introdusere en lite ønskelig ulikhet mellom folk.

SSB mener likevel at tanken om å hente inn data direkte fra større profesjonelle utleiery, i kombinasjon med datainnsamling blant de som leier av reinte private utleiery, ikke bør legges bort. Mulighetene for å benytte allerede tilgjengelige data, eller data som forholdsvis enkelt kan registreres, bør utredes videre. Vi mener det er viktig at vi går ut til så få individer som mulig for å få de opplysningene vi trenger.

2.3 Screening

Uansett hvilken metode vi velger for å etablere utvalget til en ny husleieundersøkelse, må vi gjennomføre en screening. En screening innebærer at vi går ut med et skjema til bruttoutvalget for å samle inn en del opplysninger som er nødvendige for å kunne lage en god statistikk. Det aller viktigste er å få tak i opplysninger om hvem som er leietagere slik at vi kan etablere utvalget som skal være utgangspunkt for kvartalsundersøkelsene, men det er også viktig å få informasjon om kjennetegn ved boligene og leietagerne som kan gi opphav til forskjeller i husleiene.

I FOB2001 ble det samlet inn en del opplysninger om utleieboligene som kan ha betydning for husleiene, blant annet antall rom, type rom, type bolig og alder på bolig. Dette er opplysninger som pr. i dag ikke blir oppdatert jevnlig, og som vi derfor må spørre intervjuobjektene om gjennom en årlig screening.

I etableringen av utvalget i fase 2 (se kapittel 1.1), bør det vurderes å gjennomføre en egen studie av faktorer som gir opphav til forskjeller i husleiene. Vi har nå en unik mulighet til å teste ut ulike faktorer på et forholdsvis stort utvalg, noe som vi ikke har kunnet gjøre tidligere. Dette vil medføre en noe større kostnad i fase 2, samt en mulig forskyvning i tid, men det kan bidra til at vi får en vesentlig bedre statistikk på lenger sikt.

3 Utvalgsstørrelse

Ved beregninger og anbefalinger på optimal utvalgsstørrelse må sentrale krav fra KRD ligge til grunn. For at en ny husleieindeks skal kunne bli et nyttig hjelpemiddel ved anvendelse av husleielovens bestemmelser om fastsetting og regulering av husleiene, må den kunne dekomponeres i del-indekser som viser utviklingen på regionale og lokale delmarkeder innen ulike segmenter av utleiemarkedene. Hvor stor presisjon som kreves i dataene og hvor små områder man skal produsere del-indekser for, har stor betydning for anbefalt utvalgsstørrelse.

3.1 Nivåfall vs. endringstall

Nivåfall er som regel klart mindre usikre enn endringstall. Det er også opplagt at endringstall blir svært usikre når den målte endringen er liten. Dersom en skal stille opp krav om presisjonen til endringstall må en knytte dette til hvor nøyaktige endringstall av en viss størrelse må være, f.eks. at en endring på 2 prosent kan fastslås med en nøyaktighet på 1 prosent. Det betyr i så fall at dersom endringen er målt til 2,5 prosent kan vi påstå at endringen med stor sikkerhet ligger mellom 1,5 og 3,5 prosent. For nivåfallene vil presisjonen ofte være uavhengig av nivået eller i hvert fall variere lite når nivået i de forskjellige gruppene ikke varierer for mye.

3.2 Variasjon og usikkerhet

For å svare på hvor stort utvalg som må til for å lage husleiestatistikk (indeks) må vi vite hvor stor variasjon det er i data som samles inn til formålet (månedlig eller kvartal husleie) og hvilken presisjon brukerne krever i resultatene enten det er nivåfall (gjennomsnittlig husleie) eller endringstall (endring i indeksen). I tabell 2 nedenfor har vi gitt en del tall som viser hvordan utvalgsstørrelsen varierer med spredningen i data og presisjonen i resultatet, i dette tilfellet nivåfall.

Tabell 2. Utvalgsstørrelsen som funksjon av variasjonskoeffisientene til data og resultat

CV (resultat)	CV (data)		
	25 %	37,5 %	50 %
0,5 %	2 500	5 625	10 000
1,0 %	625	1455	2 500
2,5 %	100	225	400
5,0 %	25	56	100

CV står for variasjonskoeffisienten som er lik standardavviket regnet i prosent av gjennomsnittet, altså et tall som forteller hvor stor usikkerheten er i prosent av resultatet eller gjennomsnittet av de data som er samlet inn. Det er realistisk ut fra de data vi har sett på å regne med at spredningen i data ligger imellom 25 og 50 prosent, mens et rimelig krav til resultatet kan være i området 0,5 til 5 prosent. Det virker mulig å lage en så god stratifisering av boligene(leietakerne) slik at variasjonen i leiesatsene ligger rundt 37,5 prosent innen det enkelte strata. Da ser vi at dersom en forlanger en presisjon på 0,5 prosent for landstallet krever det en utvalgsstørrelse på 5 625 boliger. Tall for mindre områder eller hustyper kan antakelig leve med en presisjon på 2,5 prosent. Tabell 2 viser at det er behov for 225 leiesatser for å oppnå dette resultatet. Gitt at dette er et akseptabelt presisjonsnivå, vil det være mulig å beregne gjennomsnittlig nivå for leiesatsene i de største fylkene med en totalstørrelse på 5 000 leietagere i alt. Med et krav til presisjon på 2,5 prosent kan det imidlertid ikke beregnes gjennomsnittlig nivå for leiesatsene i de små fylkene.

Tolkningen av en variasjonskoeffisient på henholdsvis 0,5 og 2,5 prosent er at en med 95 prosent sikkerhet kan påstå at den faktiske gjennomsnittlige leien enten ligger innenfor 1 prosent fra det publiserte landstallet og innenfor 5 prosent fra det publiserte tallet for et fylke. Tabell 2 sier derfor også noe om hvor detaljerte tall som kan publiseres i en slik undersøkelse.

I tabell 3 nedenfor har vi gjort beregninger av gjennomsnittlig husleie og variasjonskoeffisientene til husleiedataene for mars, april og desember i 2003. Dette er data som nyttes til boligindeksen i Konsumprisindeksen.

Tabell 3. Gjennomsnitt og variasjonskoeffisienten til månedsleiene for mars, april og desember 2003

Faktor	Kategori	Mars 2003			April 2003		Desember 2003	
		Antall	Gj. snitt	CV(data)	Gj. snitt	CV(data)	Gj.snitt	CV(data)
Ingen		1 222	3 880	50,0	3 942	50,9	3 976	49,6
Rom	0-2 rom	507	3 724	44,2	3 792	47,7	3 802	43,7
	3-4 rom	536	4 029	46,6	4 080	46,4	4 134	46,1
	5 + rom	179	3 873	69,9	3 954	68,9	3 995	69,9
Hustype	Enebolig og rekkehus ¹	467	3 623	57,1	3 670	57,2	3 682	56,5
	Blokk	473	4 371	45,3	4 475	46,9	4 526	44,7
	Hybel og annet	282	3 481	40,2	3 501	40,2	3 540	40,3
Fylke	02 Akershus	100	4 828	43,6	4 844	43,6	4 956	42,7
	03 Oslo	216	5 067	48,5	5 140	48,7	5 211	48,1
	07 Vestfold	58	4 115	35,6	4 149	35,4	4 176	37,7
	Gj. snitt andre fylker	53	3 389	43,0	3 443	44,0	3 469	42,6
Alle	Gjennomsnitt over strata	9,1	3 615	38,7	3 669	38,7	3 690	38,2

¹ Inkluderer også våningshus

Vi ser at variasjonskoeffisientene stort sett ligger i størrelsesorden 50 prosent eller ned mot 37,5 prosent. To tendenser i tabellen er svært påfallende:

- det er liten variasjon i gjennomsnittlig husleie etter faktorene antall rom, hustype og fylke
- det er forholdsvis liten reduksjon i variasjonen når vi stratifiserer etter antall rom, hustype og fylke - variasjonskoeffisientene reduserer fra 50% til 38-39 prosent som svarer til at effektiviteten til utvalget øker med 25 prosent

Dette kan kanskje indikere at det finnes andre faktorer som betyr mer, f.eks. alderen til leieforholdet, antallet personer i boligen(husholdningen) og boligområder. I Oslo kan avstand til sentrum og bydel være av større betydning.

3.3 Beregning av utvalgsstørrelsen

Husleiestatistikken vil omfatte både nivå tall (gjennomsnittlig husleie) og indekstall (prosentvis endring). Tallene vil bli publisert fordelt etter geografiske områder, typer boliger, husholdstyper, type kontrakt (f.eks. varighet) etc. Vi skal i dette avsnittet se litt på spørsmålet om hvor stort utvalg husleieundersøkelsen krever. Utvalgsstørrelsen styres av tre faktorer, nemlig

- hvilken presisjon krever brukeren i resultatene
- hvilken variasjon finnes i data som brukes
- hvor detaljert statistikk etterspørres

Det første og tredje ballpunktet kan behandles under ett slik at vi skal se litt på hvordan utvalgsstørrelsen kan bestemmes gitt variasjonen i data og presisjonen i data. For å si noe mer konkret må vi først innføre statistiske modeller som kan utnyttes til å beskrive sammenhengen mellom data og statistikken som skal produseres.

3.3.2 Notasjon

Vi må først innføre litt notasjon:

- h - inndeling i grupper av boliger det skal publiseres statistikk for
- s - inndeling i strata for å gjennomføre beregningene
- $Y_{i,h,s,t}$ - husleie for person i for perioden t i strata s i gruppe h
- $x_{i,h,s,t}$ - informasjon om boligen til person i for perioden t i strata s i gruppe h
- $n_{h,s,t}$ - antallet i utvalget for perioden t i stratum s i gruppe h
- $\bar{n}_{h,s,t}$ - gjennomsnittlig antall i utvalget for perioden t i strata og gruppe
- $N_{h,s,t}$ - antallet i populasjonen for perioden t i stratum s i gruppe h
- $N_{h,t}$ - antallet i populasjonen for perioden t i gruppe h
- N_t - antallet i populasjonen for perioden t

Litt mer om hva dette betyr. Forskjellen mellom h og s betyr at h kan stå for forskjellige typer husholdninger/boliger som det skal publiseres statistikk for, mens s kan være inndeling av boligene/leietakerne etter geografisk område og/eller aldersgruppe. Til sammen vil inndelingen etter h og s gi data der variasjonen er betydelig redusert innen hver gruppe og stratum sammenliknet med i data totalt. I tillegg må en regne med at variasjonen i husleie (per måned) også avhenger av størrelsen på boligene, målt i f.eks. antall rom eller kvadratmeter boligareal. Mens $Y_{i,h,s,t}$ betegner månedlig husleie kan $x_{i,h,s,t}$ betegne antall rom eller arealet til boligen.

3.3.3 Nivå tall

Det er aktuelt å publisere nivå tall som endringstall. En enkel modell for nivå tall i perioden t er

$$(1) \quad Y_{i,h,s,t} = \mu_{h,s,t} + \varepsilon_{i,h,s,t} \quad \text{der} \quad \text{Var}(\varepsilon_{i,h,s,t}) = \sigma_{h,s,t}^2$$

som betyr at så vel forventet verdi som variasjonen er avhengig av gruppen h, stratum s og perioden t.

I en slik modell vil de to ukjente parametrene, forventet verdi og variansen, kunne estimeres ved uttrykkene

$$(2) \quad \hat{\mu}_{h,s,t} = \bar{Y}_{h,s,t} = \frac{1}{n_{h,s,t}} \sum_{i \in \text{utv}_{h,s,t}} Y_{i,h,s,t}$$

$$(3) \quad \hat{\sigma}_{h,s,t}^2 = \frac{1}{n_{h,s,t} - 1} \sum_{i \in \text{utv}_{h,s,t}} (Y_{i,h,s,t} - \hat{\mu}_{h,s,t})^2$$

som er kjente størrelser, henholdsvis gjennomsnittet og standardavviket til månedsleiene i strata s i gruppe h.

Da kan vi gi ett uttrykk for standardfeilen til gjennomsnittlig månedsleie til stratum s i gruppe h gitt ved

$$(4) \quad \text{STD}(\hat{\mu}_{h,s,t}) = \frac{\hat{\sigma}_{h,s,t}}{\sqrt{n_{h,s,t}}}$$

som viser at når variasjonen i data avtar så øker også treffsikkerheten i gjennomsnittet, mens når utvalgsstørrelsen øker også treffsikkerheten. En oppgir også ofte usikkerheten eller treffsikkerheten ved variasjonskoeffisienten, lik standardfeilen dividert med gjennomsnittet(estimatet) multiplisert med hundre, gitt ved formelen

$$(5) \quad \text{CV}(\hat{\mu}_{h,s,t}) = 100 \frac{\hat{\sigma}_{h,s,t}}{\hat{\mu}_{h,s,t} \sqrt{n_{h,s,t}}} = \frac{100 \frac{\hat{\sigma}_{h,s,t}}{\hat{\mu}_{h,s,t}}}{\sqrt{n_{h,s,t}}} = \frac{\text{CV}(Y_{i,h,s,t})}{\sqrt{n_{h,s,t}}}$$

som viser hvordan variasjonskoeffisienten til resultatet varierer med variasjonskoeffisienten til data og utvalgsstørrelsen.

For vårt formål er imidlertid en formel for hvordan utvalgsstørrelsen varierer med variasjonskoeffisientene til resultatet og data viktigere.

$$(6) \quad n_{h,s,t} = \frac{\text{CV}^2(Y_{i,h,s,t})}{\text{CV}^2(\hat{\mu}_{h,s,t})} = \frac{\text{CV}^2(\text{data})}{\text{CV}^2(\text{resultat})}$$

I tabell 4 nedenfor har vi beregnet utvalgsstørrelsen for forskjellige verdier av variasjonskoeffisientene til data og resultat. Vi ser at antallet i utvalget øker med variasjonskoeffisienten til data og avtar med variasjonskoeffisienten til resultatet. Liten spredning i data og svakt krav til nøyaktigheten i resultatet krever et lite utvalg, mens stor spredning i data og strengt krav til presisjon i resultatet krever et stort utvalg. Det er rimelig å anta at variasjonskoeffisienten til data i et stratum innen en gruppe er i

området 25 til 50 prosent. Hvis vi forlanger at nøyaktigheten til resultatet må ligge i området 2,5 til 5 prosent betyr det at utvalgstørrelsen i et stratum innen en gruppe ligger mellom 25 og 400.

Tabell 4. Utvalgstørrelsen som funksjon av variasjonskoeffisientene til data og resultat

CV(resultat)	CV(data)				
	10 %	25 %	37,5 %	50 %	100 %
0,5 %	400	2 500	5 625	10 000	40 000
1,0 %	100	625	1 455	2 500	10 000
2,5 %	16	100	225	400	1 600
5,0 %	4	25	56	100	400
10,0 %	1	6	14	25	100
15,0 %	1	3	6	10	45
20,0 %	1	2	4	6	25

3.3.4 Fra strata via gruppe til totaltall

Tabell 4 gir ikke noe fullgodt svar på hvor stort utvalg det er behov for totalt siden den egentlig er beregnet for strata innen en gruppe der beregningene i først omgang gjennomføres. For å komme videre må vi stille opp formler for resultatet på gruppenivå og totalt. De er gitt ved henholdsvis

$$(7) \quad \hat{\mu}_{h,t} = \sum_s \frac{N_{h,s,t}}{N_{h,t}} \hat{\mu}_{h,s,t}$$

for gruppenivået h og

$$(8) \quad \hat{\mu}_t = \sum_h \frac{N_{h,t}}{N_t} \hat{\mu}_{h,t}$$

Dersom vi antar at andelen leieforhold i hvert stratum er konstant innen hver gruppe og dessuten variasjonskoeffisientene til hvert stratum innen hver gruppe kan regnes som konstante kan vi stille opp en enkel formel for sammenhengen mellom variasjonskoeffisienten til gruppegjennomsnittet og stratungjennomsnittet:

$$(9) \quad CV(\text{gruppe}) = \frac{CV(\text{strata})}{\sqrt{S}} = \frac{CV(\text{data})}{\sqrt{S \cdot \bar{n}_{h,s,t}}}$$

der S er antall strata. Altså dersom det er fire strata (S=4) kan vi regne med at usikkerheten halveres, f.eks. fra 10 prosent til 5 prosent. Hvis en gruppe består av S=9 strata blir usikkerheten redusert til en tredjedel. I den siste formelen er da

- $S \cdot \bar{n}_{h,s,t}$

lik antallet i utvalget for gruppe h.

Helt tilsvarende gjelder dersom vi regner at antallet i hver gruppe er omtrent det samme og vi dessuten antar at usikkerheten i hver gruppe kan regnes som konstant at

$$(10) \quad CV(\text{totalt}) = \frac{CV(\text{gruppe})}{\sqrt{H}} = \frac{CV(\text{strata})}{\sqrt{H \cdot S}} = \frac{CV(\text{data})}{\sqrt{H \cdot S \cdot \bar{n}_{h,s,t}}}$$

der H er antallet grupper. Hvis nå usikkerheten er på 5 prosent i gruppene (i gjennomsnitt) oppnår vi en presisjon på 1,7 % med 9 grupper. I den siste formelen lengst til høyre er da

- $H \cdot S \cdot \bar{n}_{h,s,t}$

lik antallet totalt i utvalget.

Hvis vi nå snur argumentene og krever 1 prosent nøyaktighet i sluttresultat, gjennomsnittlig husleie, og har en gruppe- og stratuminndeling som gir en variasjonskoeffisient på 50 prosent i data så krever det et totalt utvalg på

- $H \cdot S \cdot \bar{n}_{h,s,t} = 2\,500$

mens dersom vi reduserer kravet i resultatet til 2% kan vi under de samme betingelsene klarer vi oss med

- $H \cdot S \cdot \bar{n}_{h,s,t} = 625$

Denne formelle øvelsen vi nå har gjennomført viser at tabell 4 også gir oss svar på hvor stort utvalg det er behov for på landsbasis. Forutsetningen er at vi med rimelig god grunn kan forutsette at variasjonskoeffisientene i data er omtrent den samme for alle strata, dvs. at spredningene i husleiene er omtrent lik i alle strata.

3.3.5 Indeks - endringstall

Vi skal nå se på beregninger av indekstall og endringstall sammen med dette. Da er det naturlig å se på følgende statistiske modell

$$(11) \quad Y_{i,h,s,t} = I_{h,s,t} Y_{i,h,s,t_0} + \epsilon_{i,h,s,t}$$

der

$$(12) \quad \text{Var}(\epsilon_{i,h,s,t}) = \tau_{h,s,t}^2 Y_{i,h,s,t_0}$$

t_0 og t betegner henholdsvis perioden vi sammenligner med og perioden vi sist har samlet inn data for.

I denne modellen blir da parametrene estimert med uttrykkene

$$(13) \quad \hat{I}_{h,s,t} = \frac{\sum_i Y_{i,h,s,t}}{\sum_i Y_{i,h,s,t_0}}$$

og

$$(14) \quad \hat{\tau}_{h,s,t}^2 = \frac{1}{n_{h,s,t} - 1} \sum_i \frac{(Y_{i,h,s,t} - \hat{I}_{h,s,t} Y_{i,h,s,t_0})^2}{Y_{i,h,s,t_0}}$$

Variasjonskoeffisienten til indeksen er gitt ved formelen

$$(15) \quad \text{CV}(\hat{I}_{h,s,t}) = 100 \frac{\hat{\tau}_{h,s,t}}{\hat{I}_{h,s,t} \sqrt{\sum_i Y_{i,h,s,t_0}}} = 100 \frac{\hat{\tau}_{h,s,t}}{\hat{I}_{h,s,t} \sqrt{n_{h,s,t} \cdot \bar{Y}_{h,s,t_0}}}$$

Endringstallet er kanskje enn å mer interessant, hvor stor vekst eller nedgang i indeksen sammenliknet med perioden t_0 . Det betyr at det er

$$(16) \quad \hat{I}_{h,s,t} - 1$$

som skal undersøkes, eventuelt multiplisert med hundre for å få endringen i prosent. Variasjonskoeffisienten til endringen er gitt ved formelen

$$(17) \quad CV(\hat{I}_{h,s,t} - 1) = \frac{\hat{I}_{h,s,t}}{\hat{I}_{h,s,t} - 1} CV(\hat{I}_{h,s,t})$$

som viser at variasjonskoeffisienten kan bli svært stor når endringen er liten, siden nevneren da er liten.

For å få litt innsikt i hvordan usikkerheten på strata innen gruppene vil slå ut for gruppeindeksene og totalindeksen kan en gjøre beregninger etter samme opplegg som i formlene (11) til (17), men der en først tar vekk strata og deretter gruppe. Ved sammenlikne standardavviket i modellen på strata og gruppe vil en kunne danne seg et bilde av variasjonskoeffisientene for gruppene.

Tabell 5. Usikkerheten til indeksnivået og endringen - data for 2003

Faktor	Kategori	Antall	April mot mars			Desember mot mars		
			Indeks	CV(indeks)	CV(endr)	Indeks	CV(indeks)	CV(endr)
Rom	0-2 rom	507	101,8 %	0,94 %	52,5 %	101,9 %	0,65 %	34,8 %
	3-4 rom	536	101,0 %	0,58 %	58,9 %	102,2 %	0,51 %	23,5 %
	5 + rom	179	101,3 %	0,63 %	49,1 %	102,2 %	1,63 %	74,9 %
Hustype	Enebolig og rekkehus ²	467	101,0 %	0,70 %	73,5 %	101,2 %	0,59 %	51,1 %
	Blokk	473	102,1 %	0,95 %	46,4 %	103,0 %	0,72 %	24,3 %
	Hybel og annet	282	100,6 %	0,25 %	44,1 %	101,7 %	0,92 %	56,1 %
Fylke	02 Akershus	100	100,3 %	0,11 %	33,6 %	101,4 %	1,11 %	79,8 %
	03 Oslo	216	101,4 %	0,66 %	46,5 %	102,7 %	1,31 %	49,2 %
	Gj. snitt andre fylker	53	101,3 %	1,18 %	192,7 %	102,0 %	1,62 %	133,3 %
Alle	Gjennomsnitt over strata	9,1	101,3 %	1,56 %	94,5 %	101,8 %	3,49 %	139,8 %

En legger merke til tre klare tendenser på grunnlag av det datamaterialet vi har brukt:

- endringene er små, i gjennomsnitt over strata henholdsvis 1,3 % (mars til april) og 1,8% (mars til desember)
- usikkerheten til endringene er store for de enkelte strata, variasjonskoeffisientene er i gjennomsnitt 94,5 og 139,8 prosent over strata
- Endringene for de enkelte faktorene, antall rom, hustype eller fylke, er i mange tilfeller signifikante(variasjonskoeffisientene er under 50 prosent)

Med en utvalgsstørrelse på 5 000, dvs. et fire ganger større utvalg, tyder dette på at endringer på 1 til 2 prosent for landsdeler, hustyper og størrelse(antall rom) vil være signifikante.

² Inkluderer også våningshus

4 Oppsummering og anbefaling

4.1 Datakilde

I mangel av et løpende register over leietagere, må andre kilder velges når det skal etableres et utvalg over utleieboliger. Mulige kilder for å opprette et utvalg for innsamling av husleiedata kan være Folke- og bolig tellingen 2001 (FOB2001), Det sentrale personregisteret (DSF) eller registeret over grunneieendommer, adresser og bygninger (GAB). I FOB2001 finnes det et kjennetegn som kan identifisere leieboliger. Problemet med FOB2001 er at det ikke er etablert tilfredsstillende rutiner for å oppdatere bestanden av leieboliger i Norge. Gjennom DSF kan det etableres et utvalg leieboliger basert på oppdaterte opplysninger i forhold til FOB2001, men problemet er at det ikke eksisterer et kjennetegn på hvem som leier og hvem som eier i dette registret. Det må derfor trekkes et relativt stort bruttoutvalg som så reduseres til et nettoutvalg bestående av kun leieboliger. GAB gir opplysninger om tinglyste bolig- og eiendomskjøp. Ved å utnytte leilighetsnummer som identifiserer den som bor i en bolig (registrert bosatt), kan vi gjennom GAB trekke ut informasjon om boligen eies av den eller de som bor i boligen. Dersom den eller de som bor i boligen ifølge GAB ikke er eier av boligen, kan vi anta at den/de er leietagere. Et problem er at ikke alle boligkjøp tinglyses (f.eks. andelsboliger). Et annet problem er avviket mellom formell bolig (registrert bosatt på) og faktisk bolig (der en har sin hovedsakelige døgnhvile).

En tilnærming for å redusere utgiftene til datafangst kan være å beregne husleienivå og utvikling separat for ulike utleiegrupper som studenter, sosial boligleie gjennom kommune og utleie fra profesjonelle utleiere. Husleiedata for alle andre kategorier enn reint private utleiere hentes da fra ulike kilder som studentsamskipnaden, fra kommuner, internett og fra utleiefirma. Dette innebærer at nettoutvalget kan begrenses til rene private utleiere. Denne tilnærmingen er imidlertid heftet med en del problem som bla. å opprette ensartede og adekvate data i alle delmarkeder, overlapp mellom utleiekategoriene, ikke en uttømmende oversikt over utleiekategoriene og dermed ikke fullstendig vektinformasjon for å vekte de ulike kategoriene sammen.

Uavhengig av hvilken kilde som benyttes ved etableringen av utvalget er det nødvendig å gjennomføre en screening. Gjennom screeningen etableres det endelige utvalget, og vi får svar på en del opplysninger om boligene og leietagerne som er nødvendige for å kunne lage en god statistikk.

Vi anbefaler å trekke et bruttoutvalg fra DSF i kombinasjon med GAB for å identifisere hvorvidt den/de som er bosatt i boligen også er eier av boligen. Målpopulasjonen "leietaker" blir dermed alle de som ikke er bosatt i en bolig som eies ifølge GAB. Et problem med denne koplingen er selvfølgelig at målpopulasjonen ikke vil fange opp alle studenter ettersom disse ofte er bostedsregistrert hos foreldre. I tillegg bør et register over andelsboliger være tilgjengelig slik at andeleiere ikke inngår i vår målpopulasjon.

Enkelte adresser i flerbolighus mangler fremdeles identifikasjon (leilighetsnummer). Særlig i Oslo mangler det en del på fullføring av leilighetsnummer. En økt kvalitet på leilighetsnummer vil medføre reduserte kostnader til datafangst.

Tross enkelte problem med denne metoden, er den likevel å foretrekke fremfor en screening basert utelukkende på DSF. Dersom en ikke kopler mot GAB, må bruttoutvalget være langt større for å oppnå et nettoutvalg bestående av om lag 5000 leietagere.

Vi anbefaler også at mulighetene for å hente inn data direkte fra større profesjonelle utleiere (offentlige og private), i kombinasjon med datainnsamling blant de som leier av reint private utleiere, utredes videre.

4.2 Utvalg - størrelse og stratifisering

Det er nødvendig å trekke et bruttoutvalg på minimum 15 000 for å oppnå et nettoutvalg på 5 000 (gitt av oppdragsgiver). Erfaring fra tidligere screeninger tilsier at frafallet ofte er betydelig. Vi forventer at omlag 10 000 returnerer skjemaet i screeningen. Antar vi at om lag 2/3 av disse er leietagere, og at vi får et frafall på rundt 20 prosent på selve kvartalsundersøkelsen, får vi et nettoutvalg på i overkant av 5000. Tidligere erfaringer med screening har vist at telefonpurring reduserer frafallet betydelig. Ved telefonpurring er det derfor grunn til å anta at vi kan oppnå en svarprosent på opptil 80 prosent i screeningen. Telefonpurring er imidlertid svært ressurskrevende og neppe tilrådelig innenfor den ressursrammen som er gitt fra oppdragsgiver.

Det anbefales å stratifisere utvalget etter geografiske områder, størrelse på bolig og hustype. Som størrelsesvariabel anbefales antall rom fordi antall rom gir en mer funksjonell inndeling enn boligareal målt i kvadratmeter.

Et nettoutvalg på 5000 er såpass lite at en geografisk inndeling etter fylke med liten populasjon av leietagere kan gi svært usikre tall. Det bør derfor vurderes om landsdeler f.eks. inndelt i tettbygd ikke-tettbygd strøk er en mer hensiktsmessig geografisk stratifisering.

Forutsatt at det fremdeles er ønskelig for oppdragsgiver å vurdere husleienivå og husleieendring i og eventuelt innen de store byene, anbefales det å trekke ut byene som egne områder. For Oslo (og eventuelt de andre storbyer) bør det muligens i tillegg stratifiseres etter byregioner (sammenslåtte bydeler). En slik geografisk inndeling krever imidlertid et tilleggssutvalg for storbyene av en slik størrelse (inntil 2 500) for at det kan estimeres signifikante forskjeller i husleienivå og -endringer.

For vekting av de ulike strataene til en samlet husleieindeks, er det behov for vektinformasjon. Foreløpig har vi kun informasjon fra FOB2001 til å foreta en slik vekting. Boliger i GAB identifisert med leilighetsnummer inneholder også kjennetegn som antall rom, kvadratmeter og hustype, men ingen kjennetegn som kan identifisere om det er leietaker tilknyttet til leilighetsnummeret. Dersom bestanden av boliger som til enhver tid leies ut er noenlunde konstant med hensyn på størrelse og hustype, utgjør ikke bruk av statiske vekter fra FOB2001 et stort problem. Er de boliger som til enhver til leies ut skiftende mht. slike kjennetegn, er det derimot ønskelig med løpende (årlig) vektinformasjon.

4.3 Datafangst og frafall

Det anbefales å etablere nettoutvalget gjennom en skjemabasert screening av bruttoutvalget for å få etablert nettoutvalget og hente inn nødvendig informasjon om bolig og leietaker. Deretter følges utvalget hvert kvartal (gitt av oppdragsgiver) gjennom en CAI-basert datafangst.

Et problem som må avklares i fase 2 er behandlingen av frafall. Det er to typer frafall som preger dagens husleieundersøkelse; løpende og permanent frafall. Analyse av husleiedata for 2003 viser at det løpende frafallet ligger mellom 100 og 300 leietagere per måned med unntak av januar hvor det er noe høyere. En viktig frafallsårsak er de som nekter å være med i undersøkelsen. Analyse av nekting blant leietagerne i 2003 viser at det er klart størst andel i januar som er første gang beboerne kontaktes via telefon, men også hele 1. kvartal av 2003 viser noe flere nekte enn hva som er tilfellet i de neste kvartalene.

Det permanente frafallet oppstår når beboer som kontaktes gjennom telefonintervjuet viser seg å ikke være leieboer. Dette kan enten være fordi beboer i utgangspunktet ikke var en del av målpopulasjonen og dermed feilaktig er kommet med i utvalget, eller at det har flyttet inn en ny beboer som viser seg å være eier av boligen og dermed permanent må utgå av utvalget. I dagens husleieundersøkelse skal ny beboer i boligen vi følger spores opp når den leietakeren vi opprinnelig fulgte flytter (men det tar

ofte en viss tid før ny beboer er sport opp). Analyse av husleiedata i 2003 viser at antall parkerte er høyest i januar. Ellers i 2003 ligger antall parkerte mellom 6 og 32 per måned.

Frafall medfører at utvalget reduseres dersom en ikke supplerer gjennom året. I mangel av et husleieregister, kan det kun suppleres ved å foreta en screening av den typen som nevnt i 4.1. Ettersom screening basert på DSF og GAB ikke er hensiktsmessig metode for løpende supplering av utvalget, bør nettoutvalget ta høyde for løpende og permanent frafall.

4.4 Avsluttende merknader

I etableringen av utvalget første gang (fase 2), bør det vurderes å gjennomføre en egen studie av faktorer som gir opphav til forskjeller i husleiene. Vi har nå en unik mulighet til å teste ut ulike faktorer på et forholdsvis stort utvalg, noe som ikke har vært så enkelt tidligere. Dette vil medføre en noe større kostnad i fase 2, samt en forskyvning i tid, men det kan bidra til at vi får en vesentlig bedre statistikk på lenger sikt.

I sammenheng med utredning av faktorer som gir opphav til forskjeller i husleienivåene, anbefales det også å utrede en såkalte hedonisk metode. Denne metoden innebærer at forskjellige kvalitetsegenskaper, dvs. egenskaper ved boligen og eventuelt egenskaper ved leietaker (som f.eks. familie med utleier) dersom disse har betydning for utvikling, gis en monetær verdi. En slik metode gir oss en unik mulighet til å behandle kvalitetsendringer, f.eks. når en bolig gjennomgår en oppgradering, som kan oppstå mellom kvartal. I tillegg gir denne metoden mulighet for å linke nye boliger på de eksisterende boligene i utvalget, f.eks. dersom en supplering må foretas.

5 Litteraturliste

T.M Crone, L. I. Nakamura og R. Voith (2000): *Masuring housing services inflation* , hentet fra Journal of Economic and Social Measurement 26 (2000) side 153 - 171.

Åsmund Langsether og Per Medby (2004): *Husleieindekser og husleiestatistikk*. NOVA Rapport 10/04.

Åsmund Langsether, Lars Gulbrandsen og Erling Annaniassen (2003): *Leiemarkedet og leietakernes rettsvern* NOVA Rapport 2/03.

<http://www.ssb.no/magasinet/analyse/ husleier i konsumprisindeksen>.

<http://www.ssb.no/vis/magasinet/analyse/art-2004-02-23-02.html> *leiemarkedet - kjennetegn og prisdannelse*.

De sist utgitte publikasjonene i serien Notater

- 2004/41 K. Løyland og T.O. Thoresen: En undersøkelse av den registrerte dagmammavirksomheten. 130s.
- 2004/42 T. Nygård: Kvalitetsarbeid knyttet til kvartalsvis nasjonalregnskap (KNR) Rapport fra prosjektgruppen . 130s.
- 2004/43 E. Engelién, G. Haakonsen og M. Steinnes: Støyplage i Norge. Resultater fra førstegenerasjonsmodell for beregning av antall støyutsatte og SPI. 109s.
- 2004/44 E. Wedde: Mediebruksundersøkelsen 2003. Dokumentasjonsrapport. 32s.
- 2004/45 A.S. Abrahamsen og D. Rafat: Analyser av populasjonen i UT- prosjektet - ikke-finansielle foretak. 80s.
- 2004/46 O. Villund: Yrke i sysselsettingsstatistikken. 41s.
- 2004/47 G. Daugstad og L. Østby: Datagrunnlag for storbyutvikling. Forstudie av datagrunnlag om storbyutvikling, med særlig vekt på sosioøkonomisk og demografisk informasjon. 70s.
- 2004/48 E. Wedde, A. Holmøy, S. Skaare og O. Villund: Undersøkelse om "Utbrenthet i enkelte yrker". Dokumentasjonsrapport. 62s.
- 2004/49 H.C. Hougen: Samordnet levekårsundersøkelse 2003- tverrsnittundersøkelsen. Dokumentasjonsrapport. 83s.
- 2004/50 D.Einar Sommervoll: Slutt på billigere boliger i Oslo? OBOS-leiligheters prisutvikling 1991-2002. 25s.
- 2004/51 J.H. Wang: Kvartalsvis investeringsstatistikk. industri, bergverksdrift og kraftforsyning. 74s.
- 2004/52 J. Epland og O. Haugen: Panelutvalet til inntekts- og formuesundersøkinga 1996-2001. Dokumentasjon. 24s.
- 2004/53 KOSTRA. Arbeidsgrupperapporter 2004. 227s.
- 2004/54 T.M. Normann: Samordnet levekårsundersøkelse 2001 - panelundersøkelsen. Dokumentasjonsrapport. 54s.
- 2004/55 T.M. Normann: Samordnet levekårsundersøkelse 2002 - panelundersøkelsen. Dokumentasjonsrapport. 89s.
- 2004/56 T. Guldbrandsen og A. Holmøy: Omnibusundersøkelsen april/mai 2004. Dokumentasjonsrapport. 54s.
- 2004/57 Ø. Brekke: Praktisk guide for teknisk utstyr og dataprogrammer i brukertester. 33s.
- 2004/58 K. Henriksen: Ny metode for prismåling av personbiler i konsumprisindeksen. 24s.
- 2004/59 A.S. Abrahamsen, J. Heldal, og D. Rafat: UT- Undersøkelsene i 2004 for ikke-finansielle foretak. Utvalgsplaner og utvalg til kvartals og årsundersøkelsene. 48s.
- 2004/60 Ø. Bolsgård og L.-C. Zhang: Prisindeks for engoshandel . 35s.
- 2004/61 T. Guldbrandsen og B.O. Lagerstrøm: Undersøkelse om arbeids- og boligforhold. Dokumentasjonsrapport. 27s.
- 2004/62 G. Dahl: Trygd blant innvandrere 1992-2000. 79s.
- 2004/63 A.H. Sætre og N.Buskoven: Lokalvalgundersøkelsen 2003. Dokumentasjonsrapport. 79s.
- 2004/64 Kravspesifikasjon for elektronisk innberetning, kjennemerke og filbeskrivelse for lønnsstatistikken. Oppdatert 2004. 16s.
- 2004/65 L. Østby: Innvandrere i Norge - Hvem er de, hvordan går det med dem? Del I Demografi 156s.
- 2004/66 L. Østby: Innvandrere i Norge - Hvem er de, hvordan går det med dem? Del I Levekår 154s. ISSN 0806-3745