

*Erlend Eide Bø, Randi Johannessen og
Erling Røed Larsen*

Observerte leieinntekter for næringseiendom

Betydningen av eiendommens hedoniske attributter
og geografiske plassering

Revidert versjon

Rapporter I denne serien publiseres analyser og kommenterte statistiske resultater fra ulike undersøkelser. Undersøkelser inkluderer både utvalgsundersøkelser, tellinger og registerbaserte undersøkelser.

© Statistisk sentralbyrå, mars 2010	Standardtegn i tabeller	Symbol
Ved bruk av materiale fra denne publikasjonen skal Statistisk sentralbyrå oppgis som kilde.	Tall kan ikke forekomme	.
	Oppgave mangler	..
	Oppgave mangler foreløpig	...
	Tall kan ikke offentliggjøres	:
	Null	-
ISBN 978-82-537-7807-5 Trykt versjon	Mindre enn 0,5 av den brukte enheten	0
ISBN 978-82-537-7808-2 Elektronisk versjon	Mindre enn 0,05 av den brukte enheten	0,0
ISSN 0806-2056	Foreløpig tall	*
Emne: 08.02.30	Brudd i den loddrette serien	—
Trykk: Statistisk sentralbyrå	Brudd i den vannrette serien	
	Desimaltegn	,

Forord

Vi takker Skattedirektoratet for assistanse til trekk av data, og Finansdepartementet for kommentarer og innspill til modellkonstruksjonen. Oppdraget er utført for Finansdepartementet.

Sammendrag

Vi studerer et datasett som inneholder observasjoner om leieinntekter fra nærings-eiendom og eiendommenes attributter og plassering til å estimere en sammenheng mellom observert leie og ulike kjennetegns innvirkning på leien. Datasettet er basert på skjemarapporterte leier til Skattedirektoratet, og en grovsortering ga 32 059 observasjoner å analysere, mens en finsortering (med trunkering på ulike variablers 2. og 98. prosentil) ga 28 473 observasjoner. Resultatene viser at regresjon gir relativt god forklaringskraft med R^2 i overkant av 0,30 og med statistisk signifikante koeffisientestimer og økonomisk plausible størrelser og fortegn. Kvadratmeterleien avtar med størrelse, slik at det foreligger en rabatt på leieareal, men rabatteringenens størrelse avtar med areal. Sentrale eiendommer er dyrere enn ikke-sentrale. Kontor- og forretningsbygg er dyrere enn industribygg. Større byer gir høyere leier enn mindre byer.

Abstract

We study a dataset that includes observations on rents from commercial real estate and the attributes and location of the real estate. We use the data to estimate an association between rents and attributes/location. The dataset is based on rent reports from commercial enterprises to the Norwegian IRS (Skattedirektoratet). An initial filtering yielded 32 059 observations, while a second truncation resulted in 28 473 observations. The results demonstrate a relatively high explanatory power, statistically significant estimates, and economically plausible signs and magnitudes. The square-meter rent reduces with size, so there is a size-rebate, but the rebate diminishes with size. Central locations are more expensive than non-central. Offices are more expensive than industrial buildings. Larger cities imply higher rents than smaller towns.

Innhold

Forord	3
Sammendrag	5
Abstract	6
1. Innledning	8
2. Data	9
2.1. Rådatafil	9
2.2. Identifisering av objekt og datadetaljer	9
2.3. Vurdering av visse variabler	10
3. Empirisk analyse	11
3.1. Den empiriske modellen	11
3.2. Valg av enhet	11
3.3. Behandling av datasettet	12
3.4. Alternative gruppespesifikasjoner for bygningstyper	13
3.5. Basis og alternative enhetsspesifikasjoner	14
3.6. Regressjonsresultater	14
3.7. Delvis utleide eiendommer	15
3.8. Eiendommer hvor tomten utgjør hovedfunksjon	16
4. Diskusjon	18
4.1. Kryssvalidering	18
4.2. Eksempler og anvendelser	18
4.3. Alternative spesifikasjoner	20
5. Konkluderende merknader	21
Vedlegg A: Variabelliste	22
Tabellregister	24

1. Innledning

I den moderne økonomi er tomter og bygninger viktige størrelser. For husholdninger er boligkjøp blant de aller viktigste økonomiske beslutninger husholdningen foretar. For investorer er både næringsseiendom og boliger mulige aktiva å investere i. Det betyr at boligpriser og næringsseidomspriser påvirker husholdningers og investorers oppfatning av deres økonomiske posisjon, noe som igjen påvirker aggregert etterspørsel og derved den makroøkonomiske utviklingen. Fundamentale faktorer som renten, inntektsutviklingen og kapasitetsutnyttelsen påvirker disse prisene, men blir igjen påvirket av disse prisene i et stort endogent, feedback-system. Analyser av sammenhenger mellom observerte leiepriser og observerte kjennetegn og beliggenhet er nyttig både for boliger og næringsseiendom. De bidrar til å kvalitetsforankre indekiskonstruksjonen, utarbeide spørreskjema og oppdatere statistikkproduksjonen. I denne rapporten estimerer vi sammenhengen, altså assosiasjonen, mellom observerte leieinntekter fra utleid næringsseiendom og eiendommens attributter og plassering.

Vi finner med en enkel modell en sammenheng mellom observerte kvadratmeterleier og observerte attributter og plassering av objektet. Kvadratmeterleien avtar med størrelse, slik at det foreligger en rabatt på leieareal, men rabatteringsens størrelse avtar med areal. Sentrale eiendommer er dyrere enn ikke-sentrale. Kontor- og forretningsbygg er dyrere enn industribygg. Større byer gir høyere leier enn mindre byer. Det er imidlertid ikke slik at modellen umiddelbart kan forventes å predikere godt hvilke leier som vil oppnås for objekter som ikke er utleid. Hovedgrunnen er at det kan tenkes forskjeller mellom den næringsseidomsmassen som er utleid og den næringsseidomsmassen som ikke er utleid. I den grad de to næringsseidomsmassene er homogene, vil imidlertid modellen kunne være relativt treffsikker. Dette diskuteres nedenfor.

Beskrivelsen av arbeidet er organisert som følger. Vi lager en forholdsvis detaljert framstilling av prosessen fra mottakelse av rådata til et trunkert datasett vi kan benytte. Deretter skisserer vi grovt strukturen i vårt tankeskjema og forklarer den regresjonsteknikken vi benytter. I det påfølgende trekker vi fram en rekke empiriske funn, som vi i neste seksjon diskuterer og tester robustheten til. Til slutt konkluderer vi, og inkluderer et appendiks med relevante detaljer.

2. Data

2.1. Rådatafil

Rådatafil fra Skattedirektoratet (SKD) inneholdt totalt 61649 observasjoner. Observasjonene stammer fra innleverte næringsoppgaver; såkalt skjema RF 1098. Alle observasjoner som ikke hadde landkode Norge med verdi "NO" eller "no" ble fjernet. Videre ble alle observasjoner som ikke hadde oppgitt en brutto utleieinntekt (dvs. manglende verdi) også fjernet.

Observasjoner hvor både bruttoutleieinntekt og omregnet utleieinntekt for året var lik 0 og 1 ble fjernet. I tillegg ble observasjoner hvor areal ikke var oppgitt eller areal er lik 0 fjernet. Behandlingen ga en fil med 32 059 observasjoner.

2.2. Identifisering av objekt og datadetaljer

Vi anstrengte oss for å identifisere unike identer/objekter for analyse av utleieinntekt og andre kjennetegn på dataene. Organisasjonsnummer er ikke unikt fordi det repeteres flere ganger, og den nærliggende tolkningen er at for et gitt firma med et gitt organisasjonsnummer, kan flere næringsoppgaver (skjema RF 1098) ha vært levert. Det er i tråd med skjemabeskrivelsen. Samme organisasjonsnummer har ulike adresser og gårds- og bruksnummer, noe som indikerer at de beskriver ulike eiendommer. Det er igjen i tråd med skjemabeskrivelsen siden en organisasjon med ett organisasjonsnummer kan eie flere ulike eiendommer med ulik plassering. Organisasjonsnummer som repeteres flere ganger har fått et tilhørende nummer i en egen variabel (rep_nr). Om lag 2/3 av observasjonene omfatter kun en enhet (bygning og/eller tomt) som leies ut.

Alle objekter i filen har oppgitt gårds- og bruksnummer. Det er imidlertid ikke mulig å danne unike objekt ved å kun kombinere gårdsnummer og bruksnummer, selv ikke når vi kun velger gårds- og bruksnummer innen en spesifikk kommune. Det er også flere observasjoner hvor gårdsnummer og bruksnummer er like. Vi tolker dette som at det kan finnes flere leiere i en fysisk bygning med et gitt gårds- og bruksnummer, og det kan også finnes flere eiere av en fysisk bygning med et gitt gårds- og bruksnummer.

Det eneste som gir unik ident er å kombinere organisasjonsnummer og repetisjonsnummeret. Dette vil si at et organisasjonsnummer som repeteres for eksempel 14 ganger grunnet 14 ulike eiendommer får nummer 1 – 14 i etterkant av organisasjonsnummeret. Det tolker vi som at en organisasjon oppgir samtlige leieinntekter fra en gitt fysisk bygning (eller tomt) på ett skjema og andre leieinntekter fra en annen fysisk bygning (eller tomt) på et annet skjema. Andre eiere på de samme bygningene (og tomtene) vil selv rapportere egne leieinntekter fra disse bygningene (og tomtene) og tilsvarende samle leieinntekter fra en gitt bygning (eller tomt) på ett skjema, selv om inntektene der skulle komme fra flere leietakere. Det er dermed en eiendom med en gitt eier som blir det unike objektet, selv om dette kan åpne for flere leietakere innenfor eiendommen. Vi kommer tilbake til hvordan det er mulig å trekke ut en leieinntektsstrøm fra en leietaker her, men ikke mulig å kontrollere presist at den leietakeren vi da trekker ut er den økonomisk viktigste i leieinntektsstrømmen. Det er altså essensielt at vi har in mente at det kan oppstå utfordringer i analyser der det er avgjørende å skille mellom ulike leietakere innenfor en bygning (eller tomt) av en utleier.

En sentralitetsvariabel ble koplet på kommunenummer slik at objektene kan klassifiseres etter hvor sentralt kommunen er rangert. I henhold til Statistisk sentralbyrå (SSB) sin klassifisering, vil kommuner får tildelt verdien 0, 1, 2 eller 3 avhengig av om de ligger i nærheten av eller består av tettsteder eller byer av forskjellig størrelse. Det ble også utledet en fylkesvariabel for hvert objekt med utgangspunkt kommunenummeret.

2.3. Vurdering av visse variabler

Summert årlig utleieinntekt: Det finnes en summert årlig utleieinntekt (leie_aar_sum) i datasettet. Veiledningen tilhørende skjema RF 1098 sier følgende:

Faktisk brutto utleieinntekt i året er 170 000 kroner. Eiendommen har vært leid ut i 7 måneder i 2008. $170\,000 / 7\text{ mnd} = 24\,286 \times 12\text{ mnd} = 291\,432$ som føres i kolonne II.

Ved sesongavhengige utleieforhold (dvs. der eiendommens karakter hindrer utleie i deler av året) benyttes faktisk utleieinntekt den perioden utleien har vart, dvs. ikke omregnet til årlig utleieinntekt. For disse føres da samme beløp i kolonne III og IV. Sum faktisk utleieinntekt omregnet til årlig utleieinntekt føres i post 142.

I underkant av 10 prosent av objektene har summert årlig utleieinntekt som er ulik årlig utleieinntekt (leie_aar). Blant disse igjen er det drøye en prosent av objektene hvor også summert leie og brutto utleieinntekt (br_leie) ikke er like. Hovedtyngden av de objektene som har ulik årlig utleieinntekt og summert årlig utleieinntekt har altså samme verdi på årlig utleieinntekt og brutto utleieinntekt. Forskjeller i variabler som summert utleid areal (sum_areal) og summert årlig utleieinntekt viser at det er flere enheter på en og samme adresse. Disse enhetene vil vi ikke klare å få ut som egne objekt.

For disse objektene må vi enten bruke summert utleid areal og årlig utleieinntekt og se på alle enheter på en og samme adresse som et objekt. Alternativt må vi utelate disse objektene (som utgjør i underkant av 10 % av materialet) i selve analysen, og anvende følgende metode: Fra hvert enkeltskjema tar vi med en – 1 – enhet som er ett objekt til analysen. Igjen går vurderingen på hva som skal betraktes som ett objekt; den fysiske eiendommens totale leiestrøm fra alle leietakere til en gitt utleier (av potensielt flere) eller de ulike leiestrømmene fra enkeltleietakere (hvor flere kan leie av samme utleier) til enkeltutleiere (hvor flere utleiere sammen kan eie deler av en bygning). Ved å kombinere organisasjonsnummer og repetisjonsnummer som beskrevet tidligere, vil det være selve eiendommens genererte leiestrøm mellom leietakere innenfor en gitt bygning (eller tomt) til en gitt utleier som blir lik objektet.

Utleid areal og summert utleid areal: Rundt 10 prosent av objektene har summert utleid areal (sum areal) som er ulik variabelen utleid areal (areal). Dette må tolkes dit hen at det er oppgitt utleid areal og brutto utleieinntekter for flere enheter innen samme eiendom.

I overkant av 5 prosent av objektene har oppgitt et utleid areal på 10 kvadratmeter (kvm) eller mindre (gjelder begge variablene; areal og i sum areal). Fire objekt har et areal lik eller høyere enn 1 000 000 kvm.

Se Appendiks for ytterligere detaljer.

3. Empirisk analyse

3.1. Den empiriske modellen

I denne seksjonen tar vi for oss vår empiriske modell, og tillater noen mer tekniske og utdypende detaljer. For å estimere hvilken betydning ulike kjennetegn har på leiepris benytter vi den økonometriske teknikken med log-log regresjonsanalyse, hvor vi tar logaritmen til kvadratmeterleie og logaritmen til areal og stiller opp en sammenheng mellom dem. Regresjonsteknikken minimerer summen av kvadratavviket mellom modellanslag for leie og den faktisk observerte leien, og den kan ha en tendens til å vektlegge observasjoner med store leier fordi kvadrerte avvik her blir store. Vi har imidlertid kontrollert for en slik effekt i datatrunkeringen (se nedenfor), og vi oppnår ytterligere å dempe en slik effekt når vi altså setter modellen på logaritmeform. Den modellen vi foreslår er:

$$(1) \quad \log(\text{leie}/\text{areal}) = a + b \cdot \log(\text{areal}) + c \cdot [\log(\text{areal})]^2 + D \cdot \text{bygningstype} + E \cdot \text{sentralitetsgrad} + F \cdot \text{bystørrelse} + g \cdot \text{markør for 100 kvadratmeter} + u,$$

der enkeltbokstaver angir koeffisienter vi estimerer (små bokstaver for skalarer og store bokstaver for vektorer) og der u er et klassisk støyledd med forventning null og konstant varians. Bygningstype, sentralitetsgrad og bystørrelse er altså dummyvektorer der en klasse fungerer som utgangspunkt og der en binærvariabel lik 1 vil angi avvik fra utgangspunktet og hvilken klassifisering som passer. For eiendommer der tomten utgjør hovedfunksjonen, har vi utvidet modellen i likning (1) til:

$$(2) \quad \log(\text{leie}/\text{areal}) = a + b \cdot \log(\text{areal}) + c \cdot [\log(\text{areal})]^2 + D \cdot \text{bygningstype} + E \cdot \text{sentralitetsgrad} + F \cdot \text{bystørrelse} + g \cdot \text{markør for 100 kvadratmeter} + a_2 + b_2 \cdot [\log(\text{areal}) \cdot \text{tomt}] + u,$$

der koeffisienten a_2 er ment å ta høyde for at et annet konstantledd passer for tomter og der b_2 er ment å kontrollere for at kvadratmeterprisingen kan være annerledes for tomter.

Gitt at modellen i (1) (og (2)) passer, kan man invertere funksjonen og således oppnå en modell som kan beregne en antatt passende kvadratmeterleie basert på observerte kjennetegn og plassering slik beskrevet i (3):

$$(3) \quad \text{estimert kvadratmeterleie} = \exp[a + b \cdot \log(\text{areal}) + c \cdot [\log(\text{areal})]^2 + D \cdot \text{bygningstype} + E \cdot \text{sentralitetsgrad} + F \cdot \text{bystørrelse} + g \cdot \text{markør for 100 kvadratmeter}],$$

der \exp er korthåndsnotering for eksponensialfunksjonen (et potensuttrykk med e som grunntall), som inverterer logaritme-funksjonen, og der alle koeffisienter i likning (3) nå er estimerte koeffisienter. Dermed kan likning (3) benyttes til å forsøke å anslå priser på ikke-utleide objekter. Det foreligger imidlertid flere forbehold, og de gjennomgås nedenfor, i og med at ikke-utleide objekter kan være nettopp ikke-utleide fordi de inneholder/mangler noe som utleide har/ikke har. Med andre ord kan det være forskjell på massen av utleide næringseiendom og massen av ikke-utleide næringseiendom.

3.2. Valg av enhet

Et spørsmål når det gjelder databruk er på hvilket nivå vi skal behandle observasjonene. La oss studere et stilisert eksempel. Anta at en bygård har tre lokaler, hvert leid ut til tre forskjellige bedrifter. Bygården har en eier (men kan ha flere), og lokalenes leieinntekter rapporteres på ett skjema. Skal bygården behandles fra eiersiden, som et objekt, med totalt kvadratmetertall og total leie, eller fra leiesiden, som tre objekter, med tre arealer og tre leier? Generelt sett vil det trolig være bedre å se det fra leiesiden, i og med at det dreier seg om tre økonomiske prosesser der

relasjonen utleier-leietaker er unik og hvor de to aktørene leier og eier har drevet forhandling. Analyseteknisk gir det også flere observasjoner, og dermed også mer presise estimater på sammenhengen mellom størrelse og kvadratmeterpris. Dette følger fordi et bygg med tre lokaler på 1000 kvm kan ventes å ha høyere samlet leie enn bygg med et lokale på 3000 kvm. En tredje grunn er likebehandling, siden et bygg med tre lokaler, eid av tre forskjellige eiere, vil behandles som tre forskjellige objekter uansett.

Problemet er at det ikke har vært mulig å få inkludert alle observasjonene når flere enheter rapporteres på samme skjema. Datasettet angir riktignok leie og areal for enkeltenheter, men bruker vi det får vi med 1 (den første enheten) på hvert skjema med flere enn en enhet, men mister informasjon om de andre enhetene. En annen mulighet er å bruke total leie og totalt areal som rapporteres på hvert skjema, og da får vi med informasjon om *alle* enheter og leier, men ikke spesifisert for hver enkeltenhet, med problemene nevnt i forrige avsnitt. I tillegg vil ikke nødvendigvis alle enheter på et skjema ha samme bygningstype, noe som gir problemer ved estimering av leiepriseffekt av bygningstype. En tredje mulighet er å droppe alle observasjoner med dette problemet. Men da mister vi over 2 600 observasjoner, nesten 10 prosent av datasettet. Hovedsakelig velger vi å bruke leie og areal for hver enkeltenhet, hvor vi benytter *den første* observasjon fra skjemaer med mer enn en enhet. Men vi har også sett på hvordan resultatene blir hvis vi bruker de andre enhetsspesifikasjonene.

3.3. Behandling av datasettet

Leier, arealer og alle andre variabler i datasettet er selvrapporterte, og det foreligger ingen sanksjonsmuligheter overfor eller kontrollmuligheter av aktører som rapporterer feil. Selv etter fjerning av manglende utfylte skjemaer, gjelder det å ta hensyn til tilfeldige eller manipulerede feil. I tillegg ønsker vi en modell som er representativ for normale næringseiendommer, ikke for ekstremene. Vi trunkerer derfor datasettet på areal, leie og kvadratmeterleie ved andre og nittiåttende prosentil for å fjerne de mest ekstreme verdiene. For areal ser det ikke ut som trunkeringen tar ut et tilstrekkelig antall ekstremverdier. Det er nesten 2000 observasjoner (over seks prosent av settet) med størrelse på 1 kvm eller mindre. De 2 005 observasjonene med størrelse 10 kvm eller mindre har snittpris på nesten 900 000. De 26 observasjonene mellom 11 og 15 kvadrat har en pris på ca. 80 000. De 19 mellom 13 og 15 har en leie på 22 000 (den store endringen pga. en obs på 12 kvm med nesten 1,5 mill i leie). Vi bruker kun observasjoner med areal over 12 kvadratmeter. Observasjoner med areal større enn 10 000 kvadratmeter fjernes. Etter denne trunkeringen sitter vi igjen med 29 582 observasjoner.

Trunkeringen på andre og nittiåttende prosentil av leie (leier under 32 654 og over 8 422 816 kr fjernes) reduserer datasettet til 28 799 observasjoner. Og etter trunkering av kvadratmeterleie (under 121,13 og over 610 837 kr kvadratmeteren fjernes) har vi 28 473 observasjoner. Det er en opphopning av areal på runde tall, særlig 100 kvadratmeter. Ved å kikke på et scatter-plot av areal og leie, ser det ut som det også er en del uforholdsmessig høye leier på dette punktet. Derfor legger vi på en dummy for 100 kvadratmeter i regresjonene våre.¹

Observasjonene er delt opp i bygningstyper, og 34 forskjellige typer er representert. Det er i overkant mange typer, og noen av dem er bare representert ved svært få observasjoner. Av praktiske hensyn ønsker vi ikke mer enn fem-seks grupper. Vi har testet spesifikasjoner med flere (og færre) grupper for å se etter hvor mye tilleggsinformasjon man vinner på det. For å være representative bør ikke gruppene være for små. Utformingen av grupper bygger delvis på hvilke bygningstyper det virker logisk å slå sammen (for eksempel enebolig og tomannsbolig) og delvis på

¹ Også på andre runde arealer, for eksempel 1000 kvadratmeter, er det uvanlig mange observasjoner, men dummykoeffisientestimerer er ikke statistisk signifikante.

inspeksjon av deskriptiv statistikk for de forskjellige bygningstypene. Tabell 3.1 viser gruppefordelingen vi bruker som standard. Dummyer for gruppene 1 – 5 er inkludert i våre regresjoner.

Tabell 3.1. Inndeling av bygningstyper i grupper

Gruppe	Type	Bygningskode	Antall
0	Små bolighus	110, 120, 130 og 160	4 260
1	Større bolighus	140, 150, 170, 180 og 190	5 163
2	Industribygg	210, 220, 230 og 240	4 377
3	Kontorbygg	310	2 609
4	Forretningsbygg	320	10 818
5	Annet	410, 420, 430, 440, 510, 520, 530, 610, 620, 640, 650, 660, 670, 710, 720, 730, 820 og 830	1 258

Vi benytter som nevnt ovenfor SSBs klassifisering for sentralitet. Det er naturlig å tenke seg at sentral beliggenhet, her forstått etter kommuneklassifiseringen, vil virke positivt på leiepris. Vi legger derfor inn sentralitetsvariabelen i datasettet, og lager dummyer for sentralitet. Men det er grunn til å tro at sentralitet ikke er en tilstrekkelig finmasket sortering av geografiske forhold (kommuner som Bygland, Rømskog og Stjørdal har samme sentralitet som Oslo). Derfor legger vi også på dummyer for de store² byene i landet: Oslo, Bergen, Trondheim, Stavanger, Kristiansand, Fredrikstad, Tromsø, Sandnes og Drammen. Dummykoeffisientestimatet er statistisk signifikant for alle byene. I tillegg finner vi at en dummy for Akershus gir statistisk signifikante koeffisientestimer, og med parameterverdi større enn for Kristiansand, Fredrikstad, Sandnes og Drammen. Ut i fra parameterverdiene kan storbyene Bergen, Trondheim og Stavanger danne en felles gruppe og de andre byene (pluss Akershus) en annen gruppe.

Vår hovedmodell bruker log kvadratmeterleie (logaritmen til resultatet av leie dividert på areal) som avhengig variabel. Forklaringsvariablene er log areal, log areal opphøyd i annen, fem dummyer for bygningstype, en dummy for areal 100 kvm, tre dummyer for sentralitet og dummyer for lokalisering i Oslo, storby eller by. Denne modellen gir en forklaringskraft (R^2) på 0,314, noe som må beskrives som relativt høyt gitt tverrsnittsdata med så mange observasjoner og så få forklaringsvariable. Forklaringskraften målt ved R^2 er en observator som angir den andelen av variasjonen i de observerte leiene som modellen fanger opp i forhold til den totale variasjonen i de observerte leiene. En R^2 på 0,30 viser altså at variasjonen i modellens variable kan godtgjøre 30% av variasjonen i de observerte leiene. Parameterverdiene til alle forklaringsvariablene har forventete fortegn og er statistisk signifikante. En modell med enkeltvise dummyer for byene (totalt åtte inkludert Oslo) og Akershus gir tilnærmet samme R^2 .

I tillegg har vi undersøkt om fylker eller regioner er en hensiktsmessig forklaringsvariabel ved å lage dummyer for Nord-Norge, Vestlandet, Sørlandet, og sentrale deler av Østlandet (Østfold, Vestfold, Akershus og Oslo). Dummyen for Vestlandet ga ikke statistisk signifikante koeffisientestimer (og for Nord-Norge ikke signifikant på ett-prosentsnivå). Etter inspeksjon av fylkene separat, fant vi at en regionsoppdeling ikke ga ekstra forklaringskraft når vi allerede har tatt hensyn til byer. En alternativ modell som i tillegg til dummyene introdusert tidligere, inkluderer dummyer for alle fylkene gir kun en svært svak økning i forklaringskraft.

3.4. Alternative gruppespesifikasjoner for bygningstyper

En alternativ spesifisering for grupper slår sammen alle bygninger med kode på 100-tallet i en gruppe. De fem gruppene blir altså bolig, industri, kontor, forretning og annet. Denne gruppespesifiseringen reduserer forklaringskraften marginalt i forhold til den initiale gruppespesifiseringen. Se tabell 4.3.

² Målt ved innbyggertall.

I en tredje spesifisering benytter vi bare to grupper, og disse gruppene er dannet etter parameterverdier. Små bolighus slås sammen med industri og annet for å danne en gruppe. Resterende typer er i en annen gruppe. Denne spesifiseringen gir R^2 på 0,300; lavere enn basisspesifiseringen.

I en fjerde spesifisering slår vi sammen bygningstypene i 7 grupper, basert på parameterverdiene for dummyer for de enkelte typene. Så de bygningstypene som empirisk sett ligger nærmest hverandre i leiepris ender i samme gruppe, uten tanke på om det er en logisk eller økonomisk sammenheng (for eksempel er en gruppe dannet av kontorbygg, fritidsboliger og bygninger for overnatting). Denne spesifiseringen, kombinert med sentralitets- og bydummyer, gir R^2 på 0,332. Selv om den har noe høyere forklaringskraft enn basisspesifiseringen, er det problematisk med de tilfeldige gruppefordelingene.

3.5. Basis og alternative enhetsspesifikasjoner

Som tidligere nevnt er det utfordringer knyttet til å identifisere enheter. I initiale regresjoner – som beskrevet summarisk overfor – har vi brukt *enkeltenheters* leie og areal, altså *den første rapporterte* leieinntektsstrømmen fra en gitt leietaker mellom en gitt utleier, der det finnes flere. Vi har også komplettert basisregresjonen med regresjoner der vi benyttet totaltall (av areal og leie) på hvert skjema; altså der vi summerer areal og leie for alle utleide enheter i en gitt bygning til en gitt utleier.

3.6. Regressjonsresultater

I tabell 3.2 rapporterer vi resultatene fra fire ulike definisjoner av areal og leie. I den første (kolonne 2), vår *hovedmodell*, benytter vi areal og leie til den første rapporterte på skjemaet. Det vil være leien mellom en leietaker og en utleier i en bygning. I den andre (kolonne 3), benytter vi summen av areal og summen av leiene av alt som er rapportert på skjemaet. I den tredje (kolonne 4), benytter vi kun de observasjonene hvor leien til første enhet er lik total leie. Det er altså de observasjonene hvor bygningen kun gir en leiestrøm mellom en leietaker og en utleier. I den fjerde (kolonne 5), benytter vi de observasjonene hvor leien til første leiestrøm eller ulik summen av leiestrømmer. Det er de tilfellene hvor en bygning genererer flere leiestrømmer mellom en utleier og flere leietakere. I de tilfellene bruker vi samlet (total) leieinntekt og areal. Vi noterer oss at det er forskjellig antall observasjoner mellom kolonne 2 (28 473) og 3 (28 485). Det skyldes ulikheter fra trunkeeringen.

Tabell 3.2. Regresjoner på forskjellige enhetsspesifikasjoner

	Leie og areal (hovedmodell)	Total av leie og areal	Leie er lik total leie	Leie er ulik total leie
Konstant	8,538	8,664	8,653	9,190
Ln areal	-0,544	-0,572	-0,570	-0,714
(Ln areal) ²	0,029	0,031	0,030	0,040
Gruppe 1	0,143	0,153	0,157	0,126
Gruppe 2	-0,066	-0,038	-0,021*	-0,198
Gruppe 3	0,344	0,365	0,382	0,235
Gruppe 4	0,310	0,332	0,343	0,222
Gruppe 5	0,268	0,293	0,289	0,343
Dummy 100 kvm	0,566	0,553	0,584	0,103*
Sentralitet 1	0,069	0,070	0,063	0,128
Sentralitet 2	0,104	0,113	0,106	0,165
Sentralitet 3	0,111	0,112	0,111	0,126
Dummy Oslo	0,472	0,457	0,448	0,569
Dummy storby	0,359	0,364	0,359	0,398
Dummy by	0,215	0,215	0,216	0,197
Obs.	28 473	28 485	25 549	2 936
R ²	0,314	0,327	0,328	0,287

* betyr at koeffisienten ikke er signifikant på 1 %-nivå.

La oss knytte noen kommentarer til disse resultatene:

1. Forklaringskraften er overveiende god gitt antallet observasjoner. Det er interessant at forklaringskraften i den siste regresjonstypen (kolonne lengst til høyre) har både lavest antall observasjoner og lavest forklaringskraft. Det antyder hete-

- rogenitet i gruppen der leien til første enhet er forskjellig fra summen av leiene. En nærliggende tolkning er at vi her – ved å se på summen av leiene – slår sammen resultatet av ulike økonomiske prosesser og forhandlinger som har funnet sted til ulik tid. For eksempel kan en leietaker være en liten kiosk i gateplan som framforhandlet leie nylig mens en annen leietaker kan leie en stor kontormasse som ble framforhandlet som langtidsleie for lenge siden.
2. Koeffisientestimaterne er overveiende statistisk signifikante. Det er ikke overraskende gitt datastørrelsen, men både fortegn og størrelse er økonomisk plausible. For eksempel angir negativt koeffisientestimat på logaritmen av areal at det foreligger en slags kvantumsrabatt. Jo større areal en leietaker leier – alt annet likt (herunder bygningstype) – så vil leie per kvadratmeter falle. Men avslaget blir relativt mindre desto større arealet er, derav det positive koeffisientestimatet på andreordensleddet i logaritmen til areal.
 3. Bygningsgruppe 3 er dyrest. Det er kontorbygg. Forretningsbygg er nest dyrest. Industri er billigst. Dette lyder plausibelt gitt intuisjon på krav til bygning, inn klima, fasiliteter, beliggenhet innenfor det geografiske stratum, størrelse og depresiering som følge av bruk. I tillegg kan det tenkes at det foreligger langtidsleierabatter som industrien nok nyter mest godt av. Dette kan vi ikke empirisk belegge, men vi finner gode teoretiske grunner til å anta at slike rabatteringer foregår, og det rapporteres om slike i andre studier.
 4. Dummyen for 100 kvadratmeter gir statistisk signifikant koeffisientestimat. Det antyder at en rekke rapporter for leieobjekter som er større enn 100 kvadratmeter, rapporteres som om de var 100 kvadratmeter. Det lyder ikke urimelig, og det er en støykilde.
 5. Stigende sentralitetsgrad gir stigende leie. Det er rimelig.
 6. Større befolkningkonsentrasjon gir høyere leie, i og med at Oslo er dyrest og så faller leieprisen for bysentraenes størrelse.
 7. Det er liten forskjell i forklaringskraft mellom de to første typene regresjon; den ene hvor vi benytter første enhets rapporterte leie og areal og den andre hvor vi benytter totalsummene. Koeffisientestimaterne er for alle praktiske formål nokså like for disse to regresjonstypene, med unntak av størrelsen på leierabatten for industribygg. Vi kan ikke tilby noen overbevisende tolkning for dette funnet, og kan derfor heller ikke bruke det til å anbefale den regresjonstypen. Faktisk foretrekker vi den første, rapportert i kolonne 2, fordi den benytter enkeltleier, ikke summer – og derfor ikke blander resultatet av ulike økonomiske prosesser.

Tabell 3.3. Regresjon med og uten objekter med areal 100 kvm

	Hovedmodell	Uten objekter på 100 kvm.
Konstant	8,538	8,366
Ln areal	-0,544	-0,480
(Ln areal) ²	0,029	0,026
Gruppe 1	0,143	0,103
Gruppe 2	-0,066	-0,177
Gruppe 3	0,344	0,261
Gruppe 4	0,310	0,204
Gruppe 5	0,268	0,168
Dummy 100 kvm	0,566	-
Sentralitet 1	0,069	0,078
Sentralitet 2	0,104	0,100
Sentralitet 3	0,111	0,109
Dummy Oslo	0,472	0,493
Dummy storby	0,359	0,347
Dummy by	0,215	0,216
Obs.	28 473	27 130
R ²	0,314	0,300

Et spørsmål er hvordan man skal håndtere overhyppigheten av observasjoner på 100 kvadratmeter, som sannsynligvis skyldes bevisst eller ubevisst feilrapportering. Hvis vi utelater en slik dummy, vil kvadratmeterleien for små eiendommer bli høyere enn den sannsynligvis er, og en slik utelatelse berører det kjente problematikken som oppstår ved utelatte variabler; med tilhørende forventningsskjevheter i estimatorene. Til nå har vi brukt en dummymarkør for å kontrollere for 100 kvadratmeter. Men det er mulig å tenke seg regresjoner hvor man utelukker disse observasjonene. Tabell 3.3 viser effekten av å utelate observasjoner med areal lik 100

kvadratmeter, i stedet for å kontrollere for dem med en dummy. Det overordnede bildet er at hovedtrekkene er intakte.

3.7. Delvis utleide eiendommer

I datasettet finnes det en variabel som indikerer om hele eller deler av eiendommen er utleid. I tillegg kan man inspisere hvorvidt eiendommen ikke var utleid gjennom hele året, ved å se om faktisk utleieinntekt er oppgitt lavere enn utleieinntekt omregnet for hele året. Det kan tenkes at eiendommene som bare var delvis utleid eller utleid gjennom deler av året har noen spesielle attributter. Vi studerer dette ved å legge inn dummyer for delvis utleide eiendommer og eiendommer utleid deler av året. Å inkludere disse dummyene endrer de andre parameterverdiene og forklaringskraften marginalt, men begge dummyene er signifikante (på ett-prosentsnivå). Dummyen for utleie deler av året har positiv koeffisientverdi, dummyen for delutleie negativ.

Vi kan ikke si noe sikkert om koeffisientverdiens nivå, men det kan tenkes at eiendommer som leies ut deler av året har et visst sesongelement som gir dem ekstra verdi når de leies ut. Eksempler kan være en strandkiosk som bare leies ut om sommeren eller en skihytte som leies ut i skisesongen. Når det gjelder eiendommer som bare delvis leies ut, kan det være utelatte (ikke-observerte), negative attributter ved eiendommen som gjør at den bare er delvis leid ut, noe som forårsaker at de som leier, betaler en lavere pris.

3.8. Eiendommer hvor tomten utgjør hovedfunksjon

For en ganske liten undergruppe av observasjonene (knappe 900) er tomten eiendommens hovedfunksjon. Disse observasjonene må behandles på en litt annen måte, i og med at de ikke inneholder bygningstype og det er grunn til å anta at de økonomiske prosessene som ligger under leieforhandlingene kan være annerledes. De kan også avvike på leielengde og konjunkturfølsomhet.

Som for hoveddatasettet trunkerer vi datasettet på areal, leie og kvadratmeterleie ved andre og nittiåttende prosentil. For areal ser det heller ikke her ut som det utelukker et tilstrekkelig antall ekstremverdier. Det er 72 observasjoner (ca åtte prosent av settet) med størrelse på 4 kvm eller mindre. Disse har en snittleie på 238 000 kr, og vi anser at disse observasjonene kan fjernes. Observasjoner over nittiåttende prosentil, med areal større enn 185 000 kvadratmeter fjernes også. Etter denne trunkeringen sitter vi igjen med 809 observasjoner.

Trunkeringen på andre og nittiåttende prosentil av årlig leie virker heller ikke helt tilfredsstillende, da den etterlater noen observasjoner med 1 og 2 kroner i årlig leie. De observasjonene fjerner vi også. Leier under 3 og over 3 555 608 kr fjernes. Det reduserer datasettet til 771 observasjoner. Trunkering av kvadratmeterleie fjerner ikke noen flere observasjoner, men det er igjen noen observasjoner med kvadratmeterleie på under 1 kr. Vi har valgt å beholde dem.

Dummyer for sentralitet og for Oslo, storby eller by legges til, som i hovedsettet. Dummyer for type tomt; parkering eller havn, legges også til. Men disse dummyene inneholder til dels svært få observasjoner; det vil si at det er få observasjoner av denne typen. En regresjon med log kvadratmeterleie som avhengig variabel, og log areal, log areal kvadrert og dummies som forklaringsvariabler gir R^2 på 0,471, men en del koeffisientestimer er ikke statistisk signifikante.

Vi har også kjørt en regresjon hvor vi inkluderer disse eiendommene i datasettet som er brukt i hovedmodellen. En dummy for eiendommer hvor tomten utgjør hovedfunksjon, og denne dummyen multiplisert med log areal er lagt til i denne regresjonen. Det første gjøres for å tillate et generelt skift (rabatt eller premie) i leienivået for eiendommer hvor tomten er hovedfunksjon. Det andre gjøres for å tillate en annen rabattering/premietillegg av størrelse når tomten er hovedfunksjon.

Parameterverdiene er ikke overveiende forskjellige i forhold til standardregresjonen, men R^2 øker. Det kan muligens tolkes som at bygningene i hoveddatasettet har maksimumsstørrelse på 10 000 kvm, mens tomtene er på opptil 185 000 kvm, slik at vi opplever en effekt fra å slå sammen to datasett med ulike gjennomsnittsnivåer. For å se etter om dette driver resultatene har vi kjørt en regresjon der alle observasjoner over 10 000 kvm (99 observasjoner) kuttes. Dette endrer ikke parameterestimaten mye, men R^2 reduseres noe, noe som støtter teorien om at tomtenes størrelse påvirker den avleste forklaringskraften. Tabell 3.4 viser hovedmodellen, og de to regresjonene med tomter.

Tabell 3.4. Regresjon med eiendommer hvor tomt utgjør hovedfunksjon

	Hovedmodell	Inkludert tomter	Inkludert tomter under 10 000 kvm
Konstant	8,538	8,494	8,502
Ln areal	-0,544	-0,533	-0,535
(Ln areal) ²	0,029	0,028	0,029
Gruppe 1	0,143	0,139	0,139
Gruppe 2	-0,066	-0,066	-0,066
Gruppe 3	0,344	0,341	0,341
Gruppe 4	0,310	0,309	0,309
Gruppe 5	0,268	0,271	0,270
Dummy 100 kvm	0,566	0,557	0,556
Sentralitet 1	0,069	0,041*	0,039*
Sentralitet 2	0,104	0,112	0,112
Sentralitet 3	0,111	0,133	0,130
Dummy Oslo	0,472	0,469	0,470
Dummy storby	0,359	0,355	0,356
Dummy by	0,215	0,220	0,220
Dummy tomt		1,930	1,956
Dumtomt*Ln areal		-0,542	-0,548
Obs.	28 473	29 244	29 145
R ²	0,314	0,476	0,438

* betyr at koeffisienten ikke er signifikant på 1 %-nivå.

4. Diskusjon

4.1. Kryssvalidering

I tidsserieanalyser er det standard å kryssvalidere for å teste ut om det kan finnes ukontrollerte underliggende variable som samvarierer i tid. Da lager man typisk partisjoner av datasettet langs tidsakser, og sammenlikner koeffisientestimater fra de to settene, samt tester hvor godt de estimerte parameterne kan predikere observasjoner i den andre partisjonen.

I tverrsnittsanalyser bekymrer man seg ikke om underliggende variable med tidsvariasjon, fordi tverrsnittet (som navnet tilsier) er foretatt på samme tid (grovt sett, altså ett år). Likevel kan en kryssvalidering avsløre i hvilken grad noen innflytelsesrike observasjoner påvirker estimatene. Vi tror disse er håndtert ved at de sterke enkeltpåvirkerne er tatt ut i trunkeringen. Likevel er det en nyttig eksersis å undersøke hva koeffisientestimaterne blir, når man deler datasettet i to og kjører regresjon på hver del. Tabell 4.1 viser at det blir noe forskjell i parameterestimaterne. Vi benytter her vår standardmodell der vi ser på første enhet fra skjemaet pluss tilleggsobservasjoner fra tomter. Legg merke til at standardmodellen altså ikke benytter samlet areal og samlet husleie, men areal og husleie til første rapporterte enhet. Fordelen ved å kun se på første registrerte enhet er at vi *unngår* å skape heterogenitet ved at noen leiestrømmer har flere kilder og andre ikke.

Forskjeller i parameterestimater kan tyde på at noen enkeltobservasjoner har større betydning for totalresultatet enn andre. Vi overraskes ikke av dette, i og med at det er en viss sannsynlighet for at noen aparte observasjoner finnes i så store datasett og med så få forklaringsvariable – og fordi regresjonsmodellen minimerer kvadratavvik, slik at observasjoner med store absoluttverdier i avviket automatisk tillegges stor betydning. Vi kunne ha identifisert disse observasjonene ved å benytte teknikken som er kjent som Cook's regresjoner, der N regresjoner kjøres for datasett på størrelse N, hvor hver regresjon har størrelse N-1 i og med at en og en observasjon utelukkes. I de regresjonene hvor de innflytelsesrike observasjonene utelates, vil koeffisientestimaterne bli annerledes enn de andre – og dette vil oppdages. Men en slik ressurskrevende eksersis virker ikke nødvendig. Resultatene er robuste mot partisjonering i to like halvdel.

Tabell 4.1. Kryssvalidering

	Fullt datasett	Halvdelen	Halvdelen to
Konstant	8,494	8,429	8,546
Ln areal	-0,533	-0,519	-0,543
(Ln areal) ²	0,028	0,028	0,029
Gruppe 1	0,139	0,145	0,134
Gruppe 2	-0,066	-0,065	-0,066
Gruppe 3	0,341	0,345	0,337
Gruppe 4	0,309	0,317	0,302
Gruppe 5	0,271	0,266	0,275
Dummy 100 kvm	0,557	0,553	0,562
Sentralitet 1	0,041*	0,060*	0,023**
Sentralitet 2	0,112	0,123	0,101
Sentralitet 3	0,133	0,136	0,129
Dummy Oslo	0,469	0,466	0,474
Dummy storby	0,355	0,338	0,372
Dummy by	0,220	0,219	0,222
Dummy tomt	1,930	2,123	1,769
Dumtomt*Ln areal	-0,542	-0,575	-0,512
Obs.	29 244	14 622	14 622
R ²	0,476	0,475	0,477

* betyr at koeffisienten ikke er signifikant på 1 %-nivå.

** betyr at koeffisienten ikke er signifikant på 10 %-nivå.

4.2. Eksempler og anvendelser

Som en uformell og helt ufullstendig test – mest for illustrasjonens skyld – har vi testet ut prediksjonsevnen av vår standardmodell til leieprisen på noen eiendommer som har vært annonsert til leie. Vi har brukt annonser som ligger på finn.no, funnet

i perioden 21.01 - 22.01.2010. Vi har prøvd å finne noe variasjon i lokalisering og bygningstype, men vi har ikke benyttet noen statistisk rigorøs metode for dette utvalget. Utvalget er foretatt anekdotisk av forfatterne, men ingen som ga merkelig utslag har vært selektert bort. Vi rapporterer under det vi fant. Det viste seg vanskelig å finne annonser med oppgitte leiepriser fra små steder og for bygningstyper annet enn kontor. Tabell 4.2 gir informasjon om de eiendommene vi har sett på, og den leien modellen vår predikerer. Siste kolonne i tabellen gir raten for predikert leie som andel av annonsert leie.

Tabell 4.2. Eksempler

	Bygnings-type	Gate	Kommune	Areal	Annonsert leie	Predikert leie	Predikert/annonsert leie
1	Kontor	Vækerøvn.	Oslo	93	1 700	1 988	1,17
2	Kontor	Hoffsvn.	Oslo	125	720	1 834	2,55
3	Lager	Grefsenvn.	Oslo	160	650	1 143	1,76
4	Lager	Industrivn.	Larvik	11 250*	400	417	1,04
5	Kontor	Kanalgt.	Larvik	12	866	2 491	2,88
6	Forretning	Kr. Walbys veg	Kongsvinger	1450	550	689	1,25
7	Kontor	Storgt.	Kongsvinger	45	1 467	1 544	1,05
8	Kontor	Ægirs vei	Kristiansand	231	900	1 220	1,36
9	Kontor	Setesdalsvn.	Kristiansand	404	1 000	1 083	1,08
10	Forretning	Vestre Strandgt.	Kristiansand	582	1 600	979	0,61
11	Forretning	Foged Falhs vei	Alstahaug	1 076	620	687	1,11
12	Pub (annet)	Jernbanegt.	Åmot	250	528	889	1,68

*Inkluderer 1 125 kvadratmeter kontor.

Annonsert leie har blitt omregnet til årlig kvadratmeterleie hvor nødvendig. Er flere arealdefinisjoner oppgitt brukes bruksareal.

For det meste treffer predikasjonene relativt godt, men de er hovedsakelig litt for høye. Vi tolker dette slik: For det første er ikke-utleide eiendommer muligens noe mindre attraktive enn allerede utleide; det kan være derfor de ikke er utleide. Annonsene inneholder første type, mens vår modeller er estimert utifra den siste type. For det andre var brorparten av annonsene uten annonsert leiepris. Dermed kan det være selv-seleksjon blant dem som inkluderer leiepris, nemlig at leieprisen anses som et salgsargument, altså at den er rabattert og skal fungere som lokkemiddel. Disse annonserte leieprisene på Finn.no er altså forventet/annonsert leie og faktisk leie vet vi kun etter leiekontrakt er inngått. For det tredje er det forskjell i tid. Vår modell er basert på 2008-tall, mens annonsene er fra 2010. Konjunktursituasjonen kan ha gitt en leierabatt i løpet av finanskrisen i 2009.

Ved noen av enkeltobservasjonene er det også spesielle trekk som muligens kan forklare avvik fra modellen. Objekt 5, som har det største avviket, har areal på bare 12 kvadratmeter. At avviket er så stort kan tyde på at modellen vår ikke passer så godt på veldig små eiendommer, som er lite representert i datasettet vårt.

Objekt 10, det eneste der vi underestimerer leien, er ifølge annonsen et butikklokale på gateplan svært sentralt i Kristiansand. Trolig foreligger det et betydelig leiepremium for et slikt eksklusivt lokale. Leien for butikklokaler er sannsynligvis svært avhengig av beliggenhet, og ettersom vi ikke har en finere geografisk lokalisering enn kommune får vi ikke tatt tilstrekkelig hensyn til dette.

For objekt 3 er den predikerte leien en god del høyere enn den annonserte. Dette objektet er et lager i en underetasje. Det kan tenkes at gruppe 2, industribygg, har en ganske stor spredning på kvalitet og type bygg, og at vi dermed får litt for høye priser på lager.

Objekt 2, et kontorlokale på Skøyen, tilbys også til mye lavere leie enn vi predikerer. Det er vanskelig å se en spesifikk grunn for det, men leieprisen i annonsen er *totalpris*, ikke kvadratmeterpris. Det er uklart om fellesrom regnes i kvadratmeter-tallet eller ikke. Hvis fellesrommene er regnet med, kan det forklare noe av avviket, men det blir spekulasjon. Tallene er hovedsakelig ment illustrativt, for å vise at

modellen fanger opp noe av strukturen og kan brukes bredt, men at det foreligger mye støy og at det kan foreligge en støykilde mellom utleide og ikke-utleide objekter.

4.3. Alternative spesifikasjoner

Vi har testet hvor sensitive funnene er for den økonometriske spesifikasjonen, men vi rapporterer av plasshensyn ikke resultatene. Hovedsakelig er funnene intakte. Et eksempel kan løftes fram. Vi testet på et tidlig stadium en forklaringsmodell som i struktur likner på den vi har presentert, men hvor vi ikke benytter logaritmiske variable. Vi fant også da at kvadratmeterleie ble rabattert ved størrelse, men at det også her var kurvatur på rabatteringen (den avtok etter hvert som arealet økte). Ulempen med en slik spesifikasjon er at regresjonsteknikken som minimerer summen av kvadratavvik vil bli sensitiv for betydelige husleier, i og med at alle avvik fra modellens predikerte leienivåer og observerte leienivåer blir svært store når de kvadreres. Modellen vektlegger altså ytterpunktene i datasettet. Vi foretrekker logaritmespesifikasjonen.

Vi undersøkte også robustheten ved klassifikasjonene av bystørrelse, regioner og bygningstype. Det kan anvendes både finere og grovere sortering. Vårt klassifikasjonssystem er et kompromiss mellom ønsket å ha en enkel modell og ønsket om å treffe presist på alle determinanter. Som et eksempel på hvordan endringer i spesifikasjoner påvirker koefisientene, viser tabell 4.3 hvordan hovedmodellen, og modellen med tomter ville sett ut hvis vi brukte gruppespesifikasjon 2 (bolig, industri (gruppe 1), kontor (gruppe 2), forretning (gruppe 3) og annet (gruppe 4)) istedenfor gruppespesifikasjonen vi har brukt.

Tabell 4.3 Regresjoner med gruppespesifikasjon 2

	Hovedmodell	Inkludert tomter	Inkludert tomter under 10 000 kvm
Konstant	8,593	8,549	8,558
Ln areal	-0,543	-0,532	-0,534
(Ln areal) ²	0,030	0,029	0,029
Gruppe 1	-0,142	-0,140	-0,140
Gruppe 2	0,264	0,264	0,264
Gruppe 3	0,233	0,234	0,234
Gruppe 4	0,193	0,198	0,197
Dummy 100 kvm	0,558	0,549	0,549
Sentralitet 1	0,075	0,046	0,043*
Sentralitet 2	0,112	0,120	0,120
Sentralitet 3	0,117	0,139	0,136
Dummy Oslo	0,503	0,499	0,500
Dummy storby	0,370	0,365	0,366
Dummy by	0,219	0,224	0,224
Dummy tomt		1,872	1,898
Dumtomt*Ln areal		-0,545	-0,550
Obs.	28 473	29 244	29 145
R ²	0,311	0,474	0,436

* betyr at koefisienten ikke er signifikant på 1 %-nivå.

5. Konkluderende merknader

Fra et stort datasett med rundt 30 000 observasjoner på leier av næringseiendom og eiendommens attributter og plassering, har vi estimert en modell som beskriver sammenhengen. Utleid areal påvirker kvadratmeterleien. Det gjør også bygnings-type, sentralitet og bystørrelse. Vi finner at det trolig foreligger en overrapportering av eiendommer på 100 kvadratmeter, og dette tar vi hensyn til ved å legge inn en dummymarkør for slike.

Vi observerer at variasjoner i våre forklaringsvariable kan forklare om lag 30 prosent av variasjonene i observert leie. Det er relativt godt gitt heterogeniteten i næringseiendomsmassen og gitt antallet observasjoner.

Vi har forsøkt å illustrere – uhøytidelig og usystematisk – i hvilken grad modellen kan treffe på annonserte leier i den ikke-utleide næringsmassen. Resultatet er blandet. Det virker som om modellen treffer noe over i pris, men trolig finnes gode grunner til dette. Massen av ikke-utleid næringseiendom kan være noe forskjellig fra massen av utleid næringseiendom. Det er en grunn til at de ikke-utleide eiendommene ikke er leid ut, og det kan eksistere ikke-observerte variable som bidrar til forskjell i leie. Det er også slik at mange annonserte objekter ikke ble annonsert med pris. Det er grunn til å tro at det mindretallet som ble annonsert med pris, ble gjort det fordi prisen var rabattert og derfor et annonseargument. Potensielt kunne vår modell treffe godt for de objektene uten pris i annonsen når pris ble avtalt kontraktfestet. Det vil videre forskning kunne avdekke, og i så fall taler det for at vår modell trolig i grove trekk kan fange opp viktige – de viktigste – mønstre.

Vedlegg A: Variabelliste

Enkelte variabler i filen er døpt om (SKDs variabelnavn er oppgitt i parentes):

O_1	=	navn;
O_18	=	orgnr; (Enhet;organisasjonsnummer)
O_31713	=	komm; (EiendomNaringNorgeKommunennummer)
O_31714	=	gardsnr; (EiendomNaringNorgeGardsnummer)
O_31715	=	bruksnr; (EiendomNaringNorgeBruksnummer)
O_31716	=	seksjonsnr; (EiendomNaringNorgeSeksjonsnummer)
O_31717	=	festenr; (EiendomNaringNorgeFestenummer)
O_32022	=	land; (EiendomLandkode)
O_32023	=	adresse; (EiendomAdresse)
O_32024	=	utleie; (EiendomUtleie)
O_32025	=	bygg; (EiendomBygningstype)
O_32026	=	areal; (EiendomBygningUtleidAreal)
O_32027	=	br_leieinnt; (EiendomBygningUtleieinntektFaktiskBrutto)
O_32028	=	leie_aar; (EiendomBygningUtleieinntektOmregnetArsinntekt)
O_32029	=	sum_areal; (EiendomBygningerUtleidArealSum)
O_32030	=	leie_aar_sum; (EiendomBygningerUtleieinntektOmregnetArsinntektSum)
O_32034	=	stip_leieinnt; (EiendomUtleieinntektStipulert)
O_32035	=	br_leieinnt_aaret; (EiendomUtleieinntektInntektsaretBrutto)
O_32036	=	fradrag; (EiendomEierkostnaderFradrag)
O_32037	=	nt_leieinnt; (EiendomUtleieinntektNetto)
O_32038	=	kalk_faktor; (EiendomUtleieinntektKalkulasjonsfaktor)
O_32039	=	leieverdi; (EiendomUtleieverdiBeregnet)
O_32040	=	likningsverdi; (EiendomLikningsverdi)
Rep_nr	=	repetisjonsnummer, dvs. antall repeteringer av samme organisasjonsnummer
Fylke	=	objektets fylke oppgitt med både nr og navn. Avledet fra kommunenummer
Objekt	=	organisasjonsnummer kombinert med repetisjonsnummer
Sentral	=	objektets sentralitet i forhold til beliggenhet i tettsteder/byer av forskjellig størrelse. Koplet til kommunenummer

Kort om aktuelle variabler:

Gardsnummer (Gardsnr): alle objekt har verdi på denne variabelen

Bruksnummer (Bruksnr): alle objekt har verdi på denne variabelen

Seksjonsnummer (seksjonnr): vel 80 prosent av objektene mangler verdi på denne variabelen

Festenummer (festenr): drøye 97 prosent av objektene mangler verdi på denne variabelen

Stipulert utleieinntekt (stip_leieinnt): 86 prosent av objektene mangler verdi på denne variabelen

Brutto utleieinntekt for inntektsåret (br_leieinntekt_aaret): alle objekt har verdi på denne variabelen

Fradrag for kostnader (fradrag): alle objekt har verdi på denne variabelen

Netto utleieinntekt (nt_leieinnt): alle objekt har verdi på denne variabelen

Beregnet utleieverdi (Leieverdi): alle objekt har verdi på denne variabelen

Bygningstype (Bygg): alle objekt har verdi på denne variabelen

Eiendommen ligningsverdi (Ligningsverdi): kun et objekt mangler verdi på denne variabelen

Utleie: kun et som objekt mangler verdi på variabelen som angir om hele eller deler av eiendommen er utleid

Sentralitet

Med sentralitet menes en kommunes geografiske beliggenhet sett i forhold til tettsteder av ulik størrelse. Tettstedene deles i tre nivåer etter folketall og tilbud av funksjoner. Tettsteder på nivå 3 er landsdelssentrene (eller et folketall på minst 50 000), nivå 2 har et folketall mellom 15 000 og 50 000, nivå 1 har et folketall mellom 5 000 og 15 000. For å beskrive de muligheter befolkningen i en kommune har for arbeidsreiser til et eller flere av disse tettstedene, har en inndelt kommunene i fire sentralitetsnivåer (0-3).

Kode	Forklaring
0	Kommuner som ikke fyller kravene til reisetid fra tettstedene (minst sentrale kommuner)
1	Kommuner som omfatter et tettsted på nivå 1 eller som ligger innenfor 45 minutters reisetid fra et slikt tettsteds sentrum (mindre sentrale kommuner)
2	Kommuner som omfatter et tettsted på nivå 2 eller som ligger innenfor 60 minutters reisetid fra et slikt tettsteds sentrum (noe sentrale kommuner)
3	Kommuner som omfatter et tettsted på nivå 3 (landsdelssenter) eller som ligger innenfor 75 minutters (for Oslo:90 minutters) reisetid fra et slikt tettsteds sentrum (sentrale kommuner)

Byer

Storby inkluderer Bergen, Trondheim og Stavanger.

By inkluderer Kristiansand, Fredrikstad, Tromsø, Sandnes, Drammen og Akershus. Det innebærer at beliggenheten til et objekt må være i 1 av 4 klasser: Oslo, storby, by eller annet (som er utgangspunktet for en eventuell dummy tilordner stedet en annen klasse).

Tabellregister

3.1. Inndeling av bygningstyper i grupper	13
3.2. Regresjoner på forskjellige enhetsspesifikasjoner	14
3.3. Regresjon med og uten objekter med areal 100 kvm.....	15
3.4. Regresjon med eiendommer hvor tomt utgjør hovedfunksjon	17
4.1. Kryssvalidering	18
4.2. Eksempler	19
4.3. Regresjoner med gruppespesifikasjon 2	20