



Ny datakilde og beregningsmetode for flyreiser i KPI

TALL

SOM FORTELLER

NOTATER / DOCUMENTS

2021 / 21

Ragnhild Nygaard, Camilla Rochlenge

I serien Notater publiseres dokumentasjon, metodebeskrivelser, modellbeskrivelser og standarder.

© Statistisk sentralbyrå
Ved bruk av materiale fra denne publikasjonen skal Statistisk sentralbyrå oppgis som kilde.

Publisert 7. juli 2021

ISBN 978-82-587-1349-1 (elektronisk)
ISSN 2535-7271 (elektronisk)

Standardtegn i tabeller	Symbol
Ikke mulig å oppgi tall Tall finnes ikke på dette tidspunktet, fordi kategorien ikke var i bruk da tallene ble samlet inn.	.
Tallgrunnlag mangler Tall er ikke kommet inn i våre databaser eller er for usikre til å publiseres.	..
Vises ikke av konfidensialitetshensyn Tall publiseres ikke for å unngå å identifisere personer eller virksomheter.	:
Desimaltegn	,

Forord

Dette notatet beskriver arbeidet som er gjort for å få på plass en ny datakilde og beregningsmetode for prisindeks for passasjertransport med fly i konsumprisindeksen (KPI). De nye prismålingene ble ferdigstilt til publisering av KPI for januar 2021, men på grunn av koronapandemien og lave reisetall er disse ikke tatt i bruk før publisering av KPI for juni 2021. I Statistisk sentralbyrå er det seksjon for prisstatistikk som har hatt det faglige ansvaret for prosjektet, mens selve datainnsamlingen som skjer gjennom regelmessige uttrekk via et tilgjengelig API er etablert ved seksjon for dataplattform. Ved seksjon for prisstatistikk er det Ragnhild Nygaard og Camilla Rochlenge som har vært ansvarlig for den nye undersøkelsen. Rannveig Aasen ved seksjon for dataplattform har etablert den tekniske løsningen.

Statistisk sentralbyrå, 31.05.2021

Lasse Sandberg

Sammendrag

Ny datakilde og beregningsmetode for delindeks for passasjertransport med fly ble ferdigstilt til publisering av konsumprisindeksen (KPI) for januar 2021, men på grunn av at koronapandemien, strenge reiserestriksjoner og dermed lave reisetall blir ikke den nye metoden tatt i bruk før publisering av KPI for juni 2021.

Statistisk sentralbyrå (SSB) og seksjon for prisstatistikk har gjennom et EU-finansiert prosjekt avsluttet i 2020, sett nærmere på mulighetene for å laste ned data gjennom API løsninger. Nedlasting av data fra et globalt tilgjengelig API som dekker bredt flybilletter både innenlands og utenlands har vist seg å fungere godt. I motsetning til å laste automatisk ned data fra flyselskapenes egne nettsider er det en fordel å få tilgang til billettpriser fra ulike aktører i ett og samme datasett. Uttrekk fra API er også normalt mer tekniske robuste løsninger. Den nye metoden gjør det mulig å ta i bruk et langt større datagrunnlag enn hva som har vært benyttet tidligere.

Den nye prisindeksen er basert på prismålinger på definerte representative søk i et tilgjengelig API. Prismålingene forsøker å imitere konsumentatferd ved å søke etter de billigste flybillettene rettet mot privatreisende på ulike definerte strekninger, på gitte avreisedager og tidspunkt, innhentet på ulike definerte bestillingstidspunkt før avreise. Dette betyr at vi sammenligner tilnærmet like flybilletter over måleperioder, som er et viktig grunnprinsipp i en konsumprisindeks.

Koronapandemien har rammet luftfarten hardt og det er fortsatt knyttet stor usikkerhet til reisevirksomhet framover. Ny metode for prismålingene av flyreiser innføres derfor kun for innenlandsreiser fra og med KPI for juni 2021. Prismålinger basert på faktiske priser for utenlandsreiser vil bli vurdert forløpende, men utgår inntil videre.

Innhold

Forord	3
Sammendrag	4
Innhold	5
1. Innledning	6
2. Ny datakilde for prismålinger av flyreiser	7
2.1. Data via API	7
2.2. Representative reiser/søk	7
2.3. Bestillings- og avreisetidspunkt	9
3. Prisbegrep	9
3.1. Konsumentpriser	9
3.2. Priser i sanntid	10
4. Indeksberregning	10
4.1. Aggregering over flyvninger og aktører	10
4.2. Aggregering over avreise- og bestillingstidspunkt	12
5. Feilkilder og usikkerhet	13
5.1. Målefeil.....	13
5.2. Koronapandemi og markedsutvikling	14
Referanser	16

1. Innledning

Statistisk sentralbyrå (SSB) og seksjon for prisstatistikk har en klar strategi om å vurdere bruk av alternative datakilder i produksjonen av konsumprisindeksen (KPI). Seksjonen har lang erfaring med bruk av transaksjonsdata blant annet fra dagligvarebransjen, og har i mange år benyttet data fra internett for blant annet priser for passasjertransport med fly (prisindeks for flyreiser) som er en del av KPI. For denne prisindeksen har det vært satt opp ulike nettskrapere som automatisk har lastet ned data som er tilgjengelig på flyselskapenes sine egne nettsider. Automatiske nedlastinger av data fra internett eller «web scraping» er en svært effektiv løsning som gjør at man kan laste ned store datamengder, men samtidig er det IT teknisk utfordrende med krav om gode vedlikeholdsrutiner.

Seksjon for prisstatistikk har gjennom et EU-finansiert prosjekt avsluttet i 2020, sett nærmere på mulighetene for å laste ned data fra tilgjengelige API. API betyr «Application Programming Interface» og kan ses som hjelpeverktøy eller et programmeringsgrensesnitt for å utveksle data. Ved bruk av et API gjør man spørringer mot en database i stedet for mot nettsider. En slik løsning er normalt mer stabil enn å laste ned data direkte fra nettsidene. En løsning som tidlig ble vurdert var å se på mulighetene for å ta i bruk API data fra det globale bookingsystemet Amadeus. Denne datakilden viste seg raskt å være lovende.

En svært viktig del av prosjektet har derfor vært å få på plass en stabil og robust datainnsamlingsløsning. For å få tilgang til slike data har det vært en rekke administrative og juridiske oppgaver knyttet til dette prosjektet. Bookingsystemet Amadeus tilbyr ulike typer API rettet mot søk og bookinger innen reiseliv. Det har vært viktig å få på plass en løsning basert på priser og prisinformasjon i sanntid. I tillegg har det vært viktig å få på plass en robust IT teknisk løsning for å kunne laste ned data automatisk fra aktuell database. Dette er nå etablert i SSBs IT-avdeling noe som bidrar til å skape robuste løsninger med tanke på framtidig vedlikehold.

En siste del av prosjektet har vært å bygge et produksjonssystem og få på plass en beregningsmetode for prisindeks for flyreiser som kan utnytte en slik datakilde basert på store datamengder.

Ny datakilde og beregningsmetode for prisindeks for flyreiser ble ferdigstilt til publisering av KPI for januar 2021, men ble satt på vent på grunn av lave reisetall forårsaket av koronapandemien. Passasjertransport med fly slik den inngår i KPI utgjør 1,0 prosent av husholdningenes totale forbruksutgift for året 2021¹. Prisene på flyreiser svinger normalt mye sammenlignet med mange andre tjenester og er derfor en prisindeks som ofte utgjør er en viktig bidragsyter til den gjennomsnittlige prisutviklingen i KPI og får mye oppmerksomhet i media, hos bransjen selv og blant andre brukere.

Formålet med dette notatet er å beskrive det arbeidet som er gjort for å ta i bruk nye prismålinger for flyreiser i KPI. Kapittel 2 omhandler den nye datakilden og representative billettsøk, kapittel 3 ser nærmere på selve prisbegrepet som brukes mens kapittel 4 redegjør for indeksberegningene. Kapittel 5 peker på mulige feilkilder og årsaker til usikkerhet rundt undersøkelsen inkludert de utfordringene vi står ovenfor på grunn av koronapandemien.

¹ Forbruksandelen bygger på tall fra Nasjonalregnskapet.

2. Ny datakilde for prismålinger av flyreiser

2.1. Data via API

Å samle inn priser fra internett via nettskrapping har de siste årene blitt en stadig mer vanlig innsamlingsmetode blant statistikkbyråer. En fordel med en slik løsning er at man på denne måten kan laste ned store mengder data på en rask og effektiv måte. Ulempen er at nettsider endrer seg over tid, gjerne uten forvarsel og det kreves normalt høy teknisk kompetanse både for å utvikle og vedlikeholde slike løsninger. Et alternativ til å laste ned priser automatisk fra nettsider er å laste ned priser via tilgjengelige API.

Fra og med publisering av KPI for juni 2021 tas i bruk data fra tilgjengelig API i prismålingene for flyreiser som betyr at SSB har tilgang til en database med tilgjengelige flyvninger fra hvor man kan gjøre spørringer. Formålet med API løsninger er å tilby kontrollert tilgang til spesielle, strukturerte data eller tjenester mot en server. API kan være offentlig tilgjengelige eller de kan være låst med en API nøkkel. Det blir stadig vanligere at ulike selskap tilbyr strukturerte databaser som man kan gjøre spørringer mot.

API som brukes i dette prosjektet er gjort tilgjengelig av Amadeus. Amadeus er et globalt distribusjonssystem og plattform for blant annet bestilling og søk av flybilletter. Amadeus tilbyr en rekke ulike API som gjør det mulig å hente ut informasjon om pris og billetter i sanntid. Amadeus er verdensomspennende og er et av de største reservasjonssystemene og er tilknyttet store flyselskap slik som blant annet Air France, British Airways, Icelandair, SAS, Qantas og mange flere. Bookingsystemet Amadeus benyttes av de fleste aktørene innenfor den norske reiselivsnæringen som reisebyråer og turoperatører, når det gjelder bestilling av flybilletter, hotellopphold, leiebil med mer.

Amadeus tilbyr tilgang til en rekke API innen reiseliv. I dette prosjektet er det API «Flight Offers Search»² som er benyttet. På deres nettsider sier Amadeus at de søker blant over 500 flyselskaper for å tilby informasjon om blant annet pris, flight nr, flyplass, selskap og bagasje. Det er ikke mulig å bestille billetter i dette API³, kun søk. Hovedformålet med dette tilgjengelige API er å svare på spørsmål slik som «hva er de de billigste flyvningene fra Madrid til Paris 1. juni?».

Å laste ned priser fra et API er en mer robust løsning enn å skulle laste ned data fra ulike flyselskap sine egne nettsider. Det vil normalt være hyppigere endringer på flyselskapenes egne nettsider enn i databasene. En annen viktig fordel ved å bruke et API som dekker markedet så bredt som Amadeus sitt system gjør, er at vi kan laste ned prisdata fra ulike selskap i stedet for å sette opp og vedlikeholde roboter rettet mot ulike selskaps egne nettsider.

SSB har utført automatiske ukentlige nedlastinger av priser fra Amadeus API over lengre tid og erfaringene er at dette er en svært stabil teknisk løsning. Mens nettsider ofte kan endres uten forvarsel vil ikke databaser endres tilsvarende. Erfaringsmessig ser vi også at utvikling av nye API løsninger og forbedringer av eksisterende API varsles i god tid slik at man har mulighet til å tilpasse sine API uttrekk føre endringer skjer.

2.2. Representative reiser/søk

En KPI har som formål å måle faktiske utsalgspriser på varer og tjenester til husholdninger, dvs. priser inkl. skatter, avgifter, eventuelle rabatter etc. Gitt dette

² <https://developers.amadeus.com/self-service/category/air/api-doc/flight-offers-search>

³ Amadeus har andre API for booking og som er mer rettet mot reisebyråer.

formålet er det naturlig å gjøre visse avgrensninger i datagrunnlaget som lastes ned fra databasen.

Data lastes ned ukentlig og omfatter priser inkludert skatter og avgifter på definerte søk eller definerte turer som gjøres via Amadeus sitt «Flight Offers Search» API. Vi søker etter å fange opp mest mulig representative reiser og bestillingsmønstre gjort av privatreisende. Utover å avgrense uttrekket til å omfatte representative reiser, er det også viktig å avgrense søket for å gjøre et mest mulig effektivt og håndterbart uttrekk.

Før søk mot databasen kan gjøres er det blitt definert noen reiser eller søk som anses som representative for privatreisende. Disse definerte reisene eller søkene er etablert etter konsultasjon med ulike eksperter innen bransjen i tillegg til informasjon fra reisevaneundersøkelser.

Uttrekket omfatter kun økonomi-billetter⁴ da dette anses som mest representativt for privatreisende. Datagrunnlaget er avgrenset til kun å dekke direkte flyvninger (non-stop flyvninger).

Datagrunnlaget er avgrenset til et sett av definerte reiser/søk i Norge og mot utlandet. Innenlandsflyvninger er definert som korte helgeturer og omfatter hovedsakelig reiser ut fra Oslo til andre større byer i Norge slik som Bergen, Trondheim og Stavanger. I tillegg dekkes flere strekninger nord i landet.

For strekningene utenlands er det definert flere reiser, enkelte er definert som korte weekendturer blant annet til København, Stockholm og London, mens andre er definert som lengre turer som strekker seg over en helg blant annet til Amsterdam, Berlin, Paris og Barcelona. Det er også definert lengre interkontinentale reiser til blant annet New York og Bangkok. Tall på antall reisende fra Avinor og informasjon fra reisevaneundersøkelser ligger til grunn for valg (og vektning) av strekninger. Vi har også tall på hva som er formål bak reisene som gjør det mulig å kunne skille mellom privatreisende og personer som reiser i jobbsammenheng.

Alle kombinasjoner av reiser på de definerte søkene er inkludert i grunnlaget. Analyser viser at avreisetidspunkt er mer avgjørende for prisnivået enn når returen finner sted. De antatt minst representative returene er slettet fra datagrunnlaget, dvs. returer på korte helgeturer som skjer i intervallet kl 05-08 på morgenen. Vi antar at få privatreisende velger disse avgangene på korte turer, mens på lengre turer inkluderer vi ikke denne avgrensingen da privatreisende står ovenfor færre tilgjengelige flyvninger og dermed ikke på samme måte kan ta hensyn til avgangstidspunktet.

I tillegg til informasjon om pris (tur-retur) i norske kroner, strekning, dato og klokkeslett for avreise og retur innhentes informasjon om flight nr, selskap, flyplass, ankomsttider, antall tilgjengelige billetter, billettype, bookingklasse, innsamlingstidspunkt og informasjon om kostnader knyttet til ekstra bagasje. Utvalget av selskaper er avgrenset til dem som flyr på de definerte strekningene.

I sum er det en betydelig økning i antall prisobservasjoner som lastes ned som inngår i de nye prismålingene. Gjennom koronapandemien har antallet prisobservasjoner riktignok variert stort. Antallet priser tilgjengelig via API har naturlig nok vært redusert på grunn av både omfattende tilbud- og etterspørselssvikt gitt statlige restriksjoner knyttet til ulike innreiseforbud og reisefraråding.

⁴ Premium økonomibilletter fjernes.

2.3. Bestillings- og avreisetidspunkt

Et grunnleggende KPI-prinsipp er forutsetningen om at varene og tjenester inngår i beregningene når de konsumeres. I mange tilfeller kan det sidestilles med kjøpstidspunktet, men spesielt for tjenester slik som flyreiser er det ofte slik at kjøpstidspunktet eller bestillingstidspunktet (blir her behandlet synonymt) kan avvike med opptil flere måneder før tjenesten faktisk blir konsumert. Spesielt viktig er det å skille mellom bestillings- og konsumtidspunkt ettersom prisene endres svært ofte. Bestillingstidspunktet angir når datainnsamlingen bør gjennomføres, mens konsumtidspunktet forteller noe om når avreisen finner sted. Et viktig KPI-prinsipp er at det er først når tjenesten konsumeres, dvs i dette tilfellet når reisen finner sted, at endringen skal registreres i KPI.

Bestillingstidspunktet er en viktig prisbestemmende faktor. Normalt ser vi at sjansen for å bestille en «lavpris» billett er større jo lengre tid det er til avreise. Samtidig ser vi at det kan også være mulig å få tak på «billig-billetter» nærmere avreise dersom det fremdeles er ledige seter og etterspørsel er lav. Både innenlands- og utenlandsreiser kan kjøpes opptil flere måneder før avreise. Ideelt sett burde vi ha hatt mer informasjon om bestillingsmønstre blant privatreisende, per i dag har vi begrenset informasjon om dette. Ettersom bestillingstidspunkt er en prisbestemmende faktor innhentes prisene mange ganger før avreise. For innenlandsreiser starter innhenting av priser 3 måneder før avreise mens for enkelte av utenlandsreisene starter datainnsamlingen hele 7 måneder før avreise. Priser innhentes fram til ca. 1 uke før avreise. Avgjørelser om hvor lang tid før avreisetidspunktet priser innhentes er tatt etter råd fra eksperter i bransjen og gitt et normalisert marked. Det er naturlig å tro at bestillingsmønstre kan endre seg under krisetider slik som under koronapandemien. En pandemi skaper stor usikkerhet, endringer i smittetall og nasjonale restriksjoner kan bety at mange reisende venter med å bestille billetter til nærmere avreise.

Når det gjelder avreisetidspunkt viser Transportøkonomisk Institutt (TØI) reisevaneundersøkelser at private reiser kanskje ikke så overraskende hovedsakelig skjer i sammenheng med helg. Vi velger derfor å følge avreisetidspunkt i forbindelse med helg i prismålingene. Dag for returreise er definert ulikt mellom de ulike strekningene. Et krav er at avreise og returreise faller inn under samme måleperiode. De fleste selskap tilbyr normalt flere avganger per dag til samme destinasjon og alle flyvninger på de definerte reisene blir lastet ned. I henhold til vanlig KPI prosedyre bør det innhentes priser over en lengre periode en kun en midtuke om prisene er kjent for å være volatile over måneden.

3. Prisbegrep

3.1. Konsumentpriser

I KPI er det faktiske utsalgspriser inkludert skatter og avgifter som skal ligge til grunn for prismålingene. Det er den prisen som husholdninger må betale som skal måles. En utfordring med prisene som lastes ned er at eksempelvis bagasjegebyr ikke nødvendigvis er en del av billettprisen. Informasjon om antall reisende som benytter seg av å kjøpe merbagasje ved flyvninger er ikke kjent og flyselskapene har til nå vært restriktive med å dele denne informasjonen. Utlegg til merbagasje bør ideelt legges til den nedlastede prisen så fremt at reisende benytter seg av merbagasje slik at prisbegrepet inneholder så langt det lar seg gjøre en faktisk sluttpris. Noen reisende betaler ekstra for bagasje og andre tilleggstjenester, for andre kan det komme et tillegg for bruk av kredittkort. Vi inkluderer gebyrer som ikke er valgfrie i prisene og vi måler endring i bagasjegebyrer.

Det har blitt gjort flere sammenligninger mellom prisene som lastes ned via API og priser som er tilgjengelig på selskapenes egne nettsider. Amadeus legger ikke på et eget gebyr på billetter som selges via dette bookingsystemet. Det er avdekket at for enkelte selskap kan det være avvik i prisene mellom prisene lastet ned via API og selskapenes egne priser på deres nettsider. Dette skyldes at ikke alle priser er automatisk tilgjengelige i API for alle selskaper. Likevel synes det å være et godt samsvar mellom prisene for de fleste selskap. «Self-Service API» som benyttes viser kun publiserte flybilletter, store reiseaktører kan ha tilgang til forhandlede priser som ikke nødvendigvis er tilgjengelig via API. Ifølge Amadeus selv kan dette være en del «last minute» rabatter som ikke inngår i prisindeksen. Dette kan ha noe å si for nivået, men ikke nødvendigvis for prisutviklingen over tid.

3.2. Priser i sanntid

Data fra Amadeus kan lastes ned fra både et test- og et «live» miljø. Nedlastinger fra testmiljøet er åpent for alle uten videre godkjenninger fra aktøren selv. Dataene i testmiljøet er såkalte «cache» data som betyr at dataene kan ses som bilder på oppdaterte sanntids-data. Det er usikkert hvor ofte disse dataene blir oppdatert i dette miljøet. Det har vært viktig for SSB å få tilgang til data i sanntid. Analyser gjort mot data i begge disse miljøene viser at prisene kan avvike nokså mye og gitt den dynamikken og raske prisendringer som skjer i dette markedet har det vært veldig viktig å bruke data i sanntid. Til prisindeks for flyreiser utnyttes data fra «live» miljøet som gjør at vi hele tiden laster ned oppdaterte tilgjengelige billetter og prisinformasjon.

4. Indeksberegning

4.1. Aggregering over flyvninger og aktører

En stor fordel med data lastet ned via API er at vi får tilgang til svært mange prisobservasjoner. Vi har også den praktiske fordelen ved at vi har data fra ulike selskaper samlet i samme datagrunnlag. En bakside med denne type data er at dette ikke er transaksjonsdata. Vi måler pris på tilgjengelig billetter, ideelt sett skulle vi ha målt pris på faktisk solgte billetter. Vi har heller ikke fullgod informasjon om når privatreisende kjøper sine billetter.

I KPI er det vanlig å sammenligne varer og tjenester av lik kvalitet over to måletidspunkter. En standardtilnærming er å bruke en fast definert vare/tjenestekurv som sammenligner priser på en vare/tjeneste tilgjengelig i en fast årlig referanseperiode (eksempelvis desember året før) og i aktuell periode (måned). Eksempel på en flyreisetjeneste er en bestemt flyvning på en bestemt avreisedag og på et bestemt klokkeslett mot tilsvarende flyvning i en annen periode. En slik «matched model» tilnærming har vært brukt fram til dag i prisindeks for flyreiser. Dette er en krevende sammenligning da dette markedet er særs dynamisk der flyvninger og tidspunkter ikke alltid lett kan sammenlignes direkte. Gjeldende prosedyre har vært at om en spesifikk flyvning ikke lengre er tilgjengelig og man ikke klarer å matche to like flyvninger har den utgåtte flyvningen blitt erstattet av en annen så nær i tid og billetttype som mulig. Metoden for å måle priser på flyreiser i KPI ble sist lagt om i 2007 for å bedre kunne ta hensyn til den dynamiske prissettingen i dette markedet gjennom hovedsakelig å utvide både antall strekninger og aktører som inngikk i prisindeksen. Selv om data etter hvert ble innhentet med nettskrapping⁵, og dermed la til rette for å bruke store datamengder, ble det valgt å bruke denne tradisjonelle KPI-metodikken ved å sammenligne og matche så like flyvninger som mulig over tid. Et slikt arbeid er

⁵ De siste årene har det også blitt innhentet priser fra selskapenes egne APIer.

vanskelig å gjøre full-automatisert og har derfor vist seg å være tidkrevende rutiner og en naturlig begrensning i forhold til utvalgsstørrelse.

Prisene som inngår i prismålingene bør være representative for hele måneden. Når prisene svinger mye slik de gjør for flyreiser, er det en fordel at det er mange priser som ligger til grunn for prismålingene. Den nye metoden gir tilgang til et langt større datamateriale hvor det er lite hensiktsmessig å sammenligne en spesifikk flyvning i de to periodene. I stedet vil et gjennomsnitt av priser basert på en rekke flyvninger som innehar bestemte egenskaper og som anses av sammenlignbar kvalitet kombineres. Dette kan betraktes som et såkalt homogent produkt (HP) eller en homogen gruppe av sammenlignbare flyvninger. Fremveksten av store datamengder tilgjengelig for bruk i KPIer har medført nasjonalt og internasjonalt en debatt rundt hvor detaljert varer og tjenester må defineres, og for å fange opp dynamikken i et marked må fremveksten av HP i fremtiden ses i lys av denne utfordringen.

Et mål med definerte HP er at disse er stabile over tid samtidig som de består av varer/tjenester som antas å være av lik kvalitet. Det blir da en avveining mellom stabilitet over tid og homogenitet. Om HP defineres for bredt vil vi kunne få en skjevhet i indeksen ved at vi ikke måler «rene» prisendringer, men endringer som er forårsaket av at tjenestene er av ulik kvalitet. Dette er imidlertid mest trolig ikke en systematisk skjevhet. Om HP defineres for smalt er det stor sjanse for at HP forsvinner og det ikke finnes en match over tid. Vi velger å bruke et geometrisk gjennomsnitt av prisene/flyvningene som faller inn under HP-definisjonen da det bedre enn et aritmetisk gjennomsnitt fanger opp substitusjonen mellom de ulike flyvningene, dvs hvordan konsumentene kan velge fritt mellom de ulike flyvningene innad i et HP.

Billettprisen avhenger blant annet av når avreisen finner sted i løpet av døgnet og uken, samt hvor lang tid i forveien man har kjøpt billetten. Slike elementer bør holdes konstant for å sikre sammenlignbarhet i prisene over tid. Det er gjennomført regresjonsanalyser på datagrunnlaget for å finne prisbestemmende faktorer og disse analysene støtter en inndeling av flyvninger i begrensede tidsintervall slik som morgen-, formiddag-, ettermiddag- og kveldsbilletter. Aggregering skjer på tvers av klokkeslett innenfor disse definerte tidsintervall. Vi antar dermed at privatreisende er indifferente mellom flyvninger som går innenfor et visst tidsintervall. Vi forsøker å imitere konsumentatferd ved at vi antar at privatreisende spør seg selv om «hva er den billigste flyvningen som kan ta meg fra A til B på dette avreisetidspunktet?». Denne antagelsen om at privatreisende er indifferente innenfor et visst tidsintervall legges til grunn for reiser innad i Norge samt for strekninger til de nordiske landene der tilbudet er som størst. For de resterende utenlandsreisene er tilbudet av flyvninger mer redusert og det er ikke grunnlag for å dele prisobservasjonene inn i like mange tidssoner av dagen. For disse strekningene tas det et geometrisk gjennomsnitt av alle prisene per avreisedag. Dette er antagelser som regelmessig må vurderes. Med økt tilbud vil det kunne være grunnlag for å inndele datagrunnlaget også for lengre utenlandsreiser inn i flere tidsintervall. Dette må vurderes på nytt når vi er tilbake til å ha mer normale luftfartstider.

Vi antar også at privatreisende er indifferent mellom aktører på samme strekning. Vi legger til grunn en atferd der privatreisende søker etter billigste billett uavhengig av selskap. Dette er et punkt som kan diskuteres. Det er ikke sikkert at privatreisende mener det er av lik kvalitet å reise med eksempelvis flyselskap 1 eller flyselskap 2, men vi velger likevel å legge til grunn denne antagelsen som også kan sies å være i tråd med en nyttemaksimerende konsument. Dette er også i

tråd med formålet bak datakilden som benyttes der dataene som tilbys er søk etter billigste billetter på gitte tidspunkter.

Det er utfordrende å skulle innføre faste vekter mellom aktørene da selskapenes markedsandeler kan variere mye mellom ulike strekninger. Faste vekter gjør det også vanskelig å skulle ta inn nye aktører løpende i prisindeksen. Aktørene blir likevel indirekte vektet gjennom antallet priser de inngår med i HP. Utfordringen er at visse selskap kan ha prisnivå som ligger over (eller under) andre selskap, og således påvirke indeksen om de forsvinner ut av indeksen. En metode som reduserer disse effektene er å bruke et ekstrem-filter som kaster ut flyvninger som avviker mye fra prisgjennomsnittet i HP. Dermed er det et «gjennomsnittsnivå» som måles innenfor et definert tidsintervall. Dette kan illustreres ved at vi tar et gjennomsnitt av alle morgen-flyvninger (mellom kl 05 og 08) på en definert avreisedag. Det er mulig at ulike flyselskaper tilbyr ulike tilleggstjenester som gjør at privatreisende ikke helt er indifferent mellom flyselskaper, men vi velger likevel en slik aggregering ettersom denne prisindeksen hovedsakelig skal dekke privatmarkedet og rene økonomibilletter. Det aktuelle API er også formet mot søk av billigste billetter som betyr at om det finnes både full fleksible billetter og mindre fleksible billetter tilgjengelige på samme tid er det de mindre fleksible billettene som blir lastet ned i uttrekket gitt at disse ligger lavere i pris. Det kan være at en slik antagelse om at privatreisende er indifferente mellom selskaper er noe mindre rimelig på lengre distanser (interkontinentale), men vi velger likevel en pragmatisk løsning og bruker samme antagelse på alle strekninger.

Flyselskapene opererer med bookingklasser der grunnprisene ikke nødvendigvis endres veldig ofte. En forbruker kan likevel oppleve store prisforskjeller fra en måned til en annen. Dette fordi det kun vil være én tilgjengelig bookingklasse per hovedkategori for en gitt avgang på et gitt tidspunkt, og hvilken bookingklasse som er tilgjengelig varierer fram mot avreise. Hvilken pris man da oppnår avhenger blant annet av tilbud og etterspørsel og på hvilket tidspunkt man bestiller reisen. I likhet med metoden som har vært brukt tidligere velger vi ikke å ta hensyn til selskapenes ulike bookingklasser i prisindeksen. Dvs bookingklasser som er mer rettet mot businessmarkedet og eventuelle grupperabatter er fjernet da vi antar at disse billettene i liten grad er tilgjengelig for privatkunder. Vi anser ikke tjenesten til å være av annen kvalitet selv om bookingklassene er ulike mellom periodene som måles. Privatkunder har ofte heller ingen informasjon om hvilken bookingklasse som billetten tilhører.

4.2. Aggregering over avreise- og bestillingstidspunkt

Det måles priser for flere avreisetidspunkt per måned og prisene innhentes på flere ulike bestillingstidspunkt før avreise. Vi velger et aritmetisk gjennomsnitt av innsamlingsukene innad i samme måned slik at vi aggregerer opp til bestillingstidspunktene 1, 2, 3, 5 og 7 måneder før avreise for de respektive strekningene. Det er mest naturlig å velge et aritmetisk gjennomsnitt ved aggregering over tid da kunder ikke kan substituere tidspunkter på samme måte som mellom flyselskap.

Det aggregeres også på tvers av reisedager innad i samme måned, dvs vi aggregerer opp til en gjennomsnittspris per måned per ulike bestillingstidspunkt. Eksempelvis blir et gjennomsnitt av priser på morgen-flyvninger innsamlet for en innenlandsstrekning 1, 2 og 3 måneder før avreise for en referanseperiode (desember året før) sammenlignet med tilsvarende gjennomsnitt i aktuell periode (måned).

Om det er slik at priser mangler for visse segment estimerer vi en pris basert på tilsvarende flyvninger slik at vi sammenligner priser med avreise på tilnærmet likt

tidspunkt på dagen og tilnærmet likt bestillingstidspunkt over tid. En fordel med å estimere priser for referanseperioden er at vi kan inkludere ulike segmenter eller HP løpende gjennom året. Om det eksempelvis ikke er målt kveldsflyvninger fra strekning A til strekning B for aktuell prisreferanseperiode, men selskap senere begynner å fly i dette tidsintervallet vil det være mulig å måle pris på dette segmentet og inkludere dette i prisindeksen fortløpende gjennom året og ikke nødvendigvis vente til vi etablerer ett nytt sett av segmenter eller HP for en ny prisreferanseperiode året etter. I normale luftfartstider antar vi at innslag at estimerte priser vil være lavt i prisindeksen. Vi ser derimot at under koronapandemien kan antallet tilgjengelige flybilletter til dels være svært lavt, særlig på utenlandsstrekningene.

Videre aggregering av billettpriser opp til priser på ulike strekninger både innenlands og utenlands skjer på tvers av bestillingstidspunkt og tid på døgnet når reisen skjer. Dette gjøres med aritmetisk gjennomsnitt. I mangel på informasjon når privatreisende bestiller sine reiser antar vi derfor en jevn bestilling over alle bestillingsmånedene. Vi har heller ingen god informasjon om når på døgnet privatreisende reiser og vi likeveier dermed de ulike tidsintervallene på dagen.

Prisrelativer per strekning vektet opp til indekser for innland og for utland med reisetall for privatreisende fra Avinor. Prisindeks for eksempelvis innenlandsstrekninger kan illustreres som følgende der $P^{t/0}$ står for prisindeks for innenlandsstrekninger og P_i^t er gjennomsnittsprisene for aktuell strekning i aktuell periode, P_i^0 er prisene for aktuell strekning i prisreferanseperiode mens w_i^b viser vekten denne strekningen har i forhold til de andre innenlandsstrekningene og stammer fra en tidligere periode:

$$P^{t/0} \text{ elementær indeks} = \sum w_i^b * \frac{P_i^t}{P_i^0} \text{ der } \sum w_i^b = 1$$

Til slutt vektet prisindekser for innland og utland til en prisindeks for passasjertransport med fly samlet. Publiseringnivået for prisindeks for passasjertransport med fly videreføres ved at prisindeksen publiseres samlet og ikke for innenlands- og utenlandsflyvninger separat.

5. Feilkilder og usikkerhet

5.1. Målefeil

Tradisjonelt er KPI-målinger basert på annonserte, tilgjengelige priser (for eksempel hyllepriser i butikk eller pris på internett) og ikke faktisk omsatte varer og tjenester. Det gjelder prismålingene i denne prisindeksen også. I stadig større grad har KPI tatt i bruk pris- og mengdedata, såkalte transaksjonsdata, for å kunne måle hva som faktisk er omsatt. API dataene er ikke transaksjonsdata. I likhet med metoden som har vært brukt tidligere, måler vi pris på tilgjengelig billetter, ikke de billettene som faktisk er solgt. En kilde til usikkerhet i indeksen er at den nedlastede prisinformasjonen ikke nødvendigvis gir oss de mest kjøpte billettene. La oss anta et tenkt tilfelle der alle privatreisende kjøper billetter til visse kveldsflyvninger på en spesifikk strekning slik at det ikke fins tilgjengelige flyvninger ved våre nedlastingstidspunkt som da vil bety at vi i stedet måler mindre representative flyvninger, eksempelvis ettermiddags-flyvninger som få faktisk har booket. Vi ser at dette kan potensielt ha visse effekter i mer ekstremtilfeller, men vi antar at de nedlastede prisene normalt er en god proxy på hva som er solgt. Dette viser også at det er viktig med ikke bare hyppige nedlastinger av data, men også med riktig starttidspunkt for datainnsamlingen for å redusere denne type

usikkerheten. Det som måles er tilgjengelige billetter og dersom de er solgt så er det til den målte prisen.

En annen usikkerhet som kan knyttes til bruk av annonserte priser er innhenting av priser på kansellerte flyvninger. Når datafangsten starter opp flere måneder før avreise, kan det være, særlig under spesielle omstendigheter slik som under koronapandemien, at det er prisdata fra flere kansellerte fly som inngår i prisindeksen. Det antas at under normale luftfartstider vil andel kansellerte flyvninger utgjøre en svært liten del av grunnlaget. Det arbeides med å få på plass løsninger som bedre kontrollerer datagrunnlaget opp mot hvilke flyvninger som faktisk har blitt gjennomført, særlig under krisetider.

Datakilden gir oss prisdata, men vi vet ikke når privatreisende kjøper sine billetter. Vi vet noe om når privatkunder reiser, men utover indikasjoner fra flyselskapene om hvor lenge før avreise billetter kjøpes, så vet vi lite om når de bestiller sine reiser. Vi bruker i utgangspunktet kun vektorer etter strekning i tillegg til å vekte prisindeksene for innland og utland til en samlet prisindeks for flyreiser. Prisindeksen må dermed bygges opp av ulike uveide gjennomsnitt for å slå sammen alle prisene. Dette bidrar til usikkerhet i indeksen. På den andre siden er antall priser kraftig utvidet fra tidligere som isolert sett er med på å redusere usikkerheten rundt endringsratene.

Tilgang til transaksjonsdata er ønskelig. En datakilde som i utgangspunktet kan gi oss dette er flyselskapene selv og SSB mottar allerede i dag data direkte fra norske selskaper til bruk i en kvartalsvis produsentprisindeks for flyreiser. Utfordringen er at en månedlig KPI som skal måle de prisene private konsumenter står ovenfor har strenge krav til aktualitet, som gjør at selskapene ikke ser at de har mulighet til å bistå oss med data i tide.

En annen utfordring er knyttet til sammenlignbarhet over tid. En KPI skal som tidligere nevnt måle prisendringer fri for endringer i kvaliteten på tjenesten som måles. Om vi har definert HP for bredt vil vi ha prisendringer forårsaket av endret kvalitet i tjenesten. Å aggregere på tvers av aktører kan eksempelvis diskuteres da flere kan si at en reise med ett flyselskap ikke er det samme som en reise med et annet selskap. Det er vurdert at dette er en rimelig antagelse på korte reiser, særlig innenlands. På lengre interkontinentale reiser kan dette være et mer reelt spørsmål, men vi velger å benytte samme beregningsmetode både på innland og utland. Et annet usikkerhetsmoment knyttet til sammenlignbarhet i tjenesten over tid er innslag av ulike billett-typer. I utgangspunktet er det viktig at vi sammenligner billett-typer av like betingelser. Full-fleksible billetter er det ofte knyttet til andre betingelser (eksempelvis kan disse være fullt refunderbare) sammenlignet med mer lavpris-billetter. Om det er slik at det kun er full-fleksible billetter som er tilgjengelige på visse avreisetidspunkt kan det være at API returnerer pris på disse billettene. Vi antar likevel at innslaget av full-fleksible billetter vil være lavt i denne prisindeksen da prisene normalt er høyere på disse billettene sammenlignet med billett-typer med mer begrensede betingelser. Om det er tilfeller der API returnerer slike full-fleksible billetter er det også sannsynlig at disse blir kastet ut på grunn av ikke-representative bookingklasser.

5.2. Koronapandemi og markedsutvikling

I 2020 ble reiselivsbransjen tilnærmet lammet av koronapandemien og luftfarten er en av de bransjene som er kraftigst berørt. Det er en betydelig bransje som berøres. I en årrekke har nordmenn vært i Europa-toppen når det gjelder antall flyturer. Rutene mellom Oslo og Bergen, Trondheim og Stavanger er blant de mest travle i Europa, og norske flyselskap er betydelige også i internasjonal sammenheng.

Tall fra Avinor viser en kraftig fall i antall passasjerer etter utbruddet av pandemien, nedgangen har vært kraftigst for utenlandsflyvningene med reisetall som har vist nedgang på om lag 90 prosent sammenlignet med tilsvarende perioder før koronautbruddet.

Dette synes også klart i API datagrunnlaget gjennom denne perioden der antallet flyvninger har vært kraftig redusert og på flere strekninger, særlig utenlandsstrekninger, faktisk falt helt bort. Det er også knyttet usikkerhet til når reisetallene igjen vil ta seg opp og normaliseres. Fortsatt pågår pandemien og det er særlig lite reising på tvers av landegrenser. Det kan ta lang tid før luftfarten normaliseres igjen.

Flere analytikere peker på at vi kan først få en normalisering av markedet i 2023 gitt at koronavaksinering går som planlagt i 2021. Vi vet heller ikke om luftfarten kan bli permanent endret av dette. Da koronakrisen oppstod var det allerede tegn til en viss flyskam drevet av miljøpolitiske faktorer, men det er uvisst hvorvidt dette har gitt og eventuelt vil gi store utslag i etterspørselen etter flyreiser. Vi vet heller ikke om koronakrisen vil bringe med seg en ny type atferd som for eksempel flyskrekk knyttet til frykt for smitte både under reisen og på reisemålet. På en annen side kan det være at reiselyst har bygd seg opp over lengre tid noe som vil kunne lede til kraftig vekst i reisetall så snart reiserestriksjonene faller bort. En permanent endring i private konsumenters reisevaner over tid vil kunne vise seg gjennom endringer i private husholdninger private konsum og dermed i den aggregerte vekten for flyreiser i KPI.

Pandemien har skapt store utfordringer for de offisielle prismålingene siden utbruddet i mars 2020. I henhold til internasjonale retningslinjer som ble utarbeidet da koronapandemien brøt ut har det vært enighet om å estimere et prisforløp for deler av konsumet som helt eller delvis har falt bort under pandemien. Et særlig fokus har det vært på komponenter med et tydelig sesongmønster i prisene da eventuelle brudd på disse kan skape stor effekt på den totale KPI veksten. Priser på utenlandsstrekninger svinger mye fra måned til måned og viser normalt et klart sesongmønster gjennom året og har derfor blitt framskrevet med sesongfaktorer helt siden mars 2020. Faktiske priser er ikke blitt brukt da en svært lav andel privatreisende har reist utenlands siden koronapandemien startet.

De historiske prismålingene i KPI for innenlandsflyvninger viser derimot ikke det samme klare sesongmønsteret gjennom året og prisene i den offisielle delindeksen ble derfor estimert basert på prisutviklingen i totalindeksen for KPI i perioden mars - juni 2020 og i perioden januar 2021 - mai 2021 da reisetallene var på sitt laveste. En slik behandling er likestilt med et bortfall av dette konsumet i KPI. I perioden juli til desember 2020 var det faktiske priser som inngikk i prisindeksen da reisetallene innad i landet tok seg noe opp. I januar 2021 ble det innført ytterligere nasjonale tiltak mot pandemien og prisene for innenlandsreiser ble igjen estimert.

Til tross for at vaksinering er godt i gang og det er nå gradvis økende reiseaktivitet innenlands, er det per mai 2021 fremdeles strenge nasjonale reiseråd knyttet til reiser til utlandet. Smittesituasjonen kan fort endre seg og det er store geografiske forskjeller i smittetrykk. Det er fortsatt bruk av reisekarantene i Norge og flere andre land. Frykten for smitte, samt reiserestriksjoner gir usikkerhet knyttet til utviklingen i reisevirksomheten utenlands. Ny metode for prismåling av flyreiser tas derfor i bruk i første omgang kun for innenlandsreiser fra og med KPI for juni. Prismålinger basert på faktiske priser for utenlandsreiser vil bli innlemmet når reisetall viser en betydelig økning, samt at de strenge reiserestriksjonene lempes på.

Referanser

Amadeus Self-Service APIs. Tilgjengelig på: <https://developers.amadeus.com/>

Avinor (2020). «Reisevaner på fly 2019». Rapport. Tilgjengelig på:

<https://avinor.no/contentassets/b3f78317709344549da26fa3ebda6977/avinor-reisevaneundersokelsen-2019.pdf>

Johansen I. (2007). «Ny metode for flyreiser i KPI». Notater 2007/40. Tilgjengelig på: https://www.ssb.no/a/publikasjoner/pdf/notat_200740/notat_200740.pdf

Statistisk sentralbyrå. «Objective F: Web scraping as a source for HICP». Final report 2017. Grant agreement 04151.2015.004-2015.629. Upublisert rapport.

Statistisk sentralbyrå. «Objective 1C: The use and analysis of prices collected on internet». Final report 2016. Grant agreement 04151.2013.001-2013.409. Upublisert rapport.

Transportøkonomisk Institutt (2018). «Reisevaner på fly 2017». TØI rapport 1646/2018. Tilgjengelig på:

https://avinor.no/globalassets/_konsern/om-oss/rapporter/reisevaneundersokelsen-2017.pdf