

Direkte kostnader

METODIKK OG BEREGNINGER AV DIREKTE KOSTNADER FOR
PRISING AV DEN MINSTE PAKKEN MED TJENESTER

Direkte kostnader

1 Innhold

1	Innhold.....	1
2	Innledning.....	3
3	Regelverk.....	3
3.1	Generelt.....	3
3.2	Differensiering av gjennomsnittlige enhetskostnader.....	3
4	Prismodell.....	3
4.1	Fremgangsmåte.....	3
4.1.1	Avgrensning av kostnadsgrunnlag.....	4
4.1.2	Valg av beregningsmetode.....	4
4.1.3	Veien videre.....	5
4.2	Data.....	6
4.2.1	Kostnadsdata.....	6
4.2.2	Tekniske data.....	6
4.2.3	Trafikkdata.....	6
4.2.4	Prisindeks.....	6
4.2.5	Datsett.....	7
4.2.6	Svakheter ved dataene.....	8
4.3	Modell og vurderinger.....	9
4.3.1	Forutsetninger for regresjonsmodellen.....	10
4.3.2	Vurdering av funksjonsform.....	11
4.3.3	Vurdering av hvilke variabler som skal inngå i modellen.....	12
4.3.4	Differensiering mellom baneområder/regioner.....	13
4.3.5	Justering av prisnivå.....	14
4.3.6	Mål på trafikken.....	14
4.3.7	Valgt modell.....	14
4.3.8	Resultater av regresjonsmodellen.....	15
5	Resultater.....	16
6	Implementering og anbefalinger.....	16
7	Bibliografi.....	17

Bane NOR

Metodikk og beregninger av direkte kostnader

8	Vedlegg A – Begreper og uttrykk.....	19
9	Vedlegg B – Regresjonsresultater ulike modeller	19
10	Vedlegg C – modell med og uten tidstrend.....	21
11	Vedlegg D – tester av modellen	22

2 Innledning

I 2021 og 2022 gjennomførte Bane NOR et prosjekt for prising av den minste pakken med tjenester¹. Denne rapporten omhandler datagrunnlag, metode, resultater og generelt hvilke vurderinger som ble gjort i beregningen av ny grunnpris². Prisene omtalt i denne rapporten vil inngå i høringen av Network Statement 2024 og vil være gjeldende fra ruteplan 2024.

I kapittel 3 kommer en kort gjennomgang av regelverket, i kapittel 4 blir prismodellen beskrevet i detalj, mens resultater presenteres i kapittel 5.

3 Regelverk

3.1 Generelt

Jernbaneforskriften § 6-2 (1) sier at «Avgiftene for bruk av tjenester nevnt i § 4-1 og for tilgang til jernbaneinfrastruktur som knytter sammen serviceanlegg, skal fastsettes til kostnaden som oppstår som en direkte følge av den enkelte togtjenesten. Avgiftene kan endres som angitt i annet og tredje ledd og § 6-3».

Jernbaneforskriften beskriver regler for både hvordan kapasitetsfordelingen og prisingen av infrastrukturtenestene som Bane NOR tilbyr, skal gjennomføres. Forskriften bygger på EU-direktiv 2012/34, og angir at prisen for den minste pakken med tjenester skal settes til de direkte kostnadene knyttet til produksjonen av tjenesten. EU har i tillegg kommet med en forordning (EU 2015/909) (forordningen) som gir nærmere føring for hvordan de direkte kostnadene skal beregnes.

Forordningen åpner i artikkel 6 for at infrastrukturforvalter kan benytte følgende metoder til å beregne direkte enhetskostnader: aktivitetsbasert kostnadskalkyle (forordningsmetoden), økonometrisk metode, teknisk kostnadsmodellering, eller en kombinasjon av disse.

I artikkel 2 (2) står det at «direkte enhetskostnad» er direkte kostnad per togkilometer, kjøretøykilometer, et togs bruttotonnkilometer eller en kombinasjon av disse.

3.2 Differensiering av gjennomsnittlige enhetskostnader

Under artikkel 5 (2) i forordning 2015/909 åpnes det for å differensiere de direkte kostnadene basert på ulike parametere som kan påvirke slitasjen på infrastrukturen. Her nevnes blant annet masse, aksellast, hastighet og toglengde m.m.

I artikkel 5 (1) i forordning 2015/909 står det at man kan ha ulikt nivå på grunnprisen på ulike deler av jernbanenettet dersom man kan godtgjøre overfor reguleringsorganet at de direkte kostnadene er svært ulike.

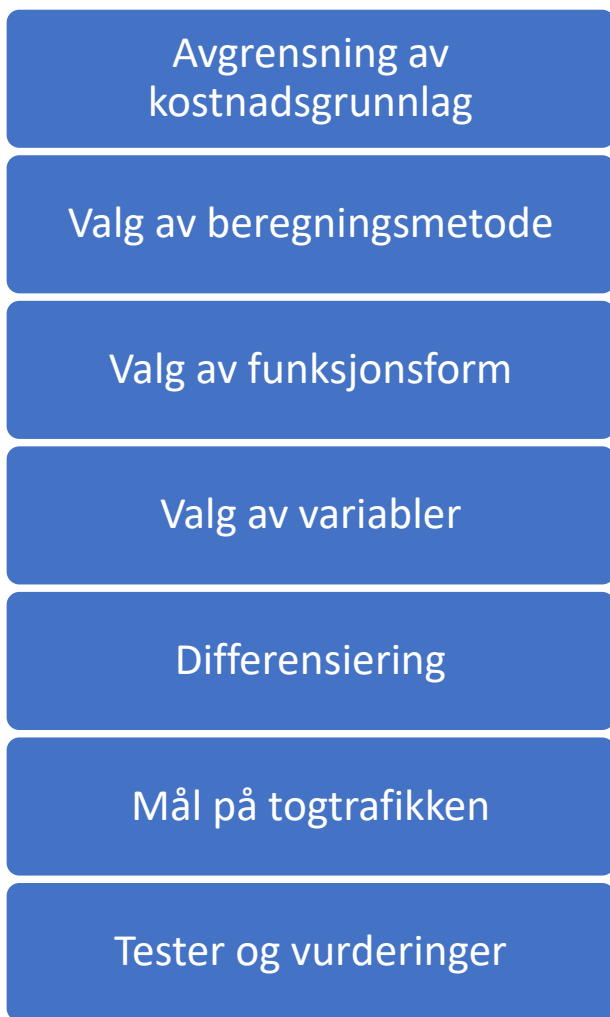
4 Prismodell

4.1 Fremgangsmåte

Beregning av de direkte trafikkavhengige kostnadene ved minstepakken gjøres i flere trinn, som illustrert i Figur 1.

¹ Den minste pakken med tjenester er definert i jernbaneforskriften § 4-1

² Pris for den minste pakken med adgangstjenester og for adgang til infrastruktur som knytter sammen serviceanlegg, i samsvar med artikkel 31 nr. 3 i direktiv 2012/34. Prisen skal fastsettes til kostnaden som oppstår som en direkte følge av driften av togtjenesten (direkte kostnad).



Figur 1 Fremgangsmåte for beregning av direkte trafikkavhengige kostnader

4.1.1 Avgrensning av kostnadsgrunnlag

De direkte kostnadene er en del av kostnadene for kjernevirksomheten. I beregningen er det kun tatt hensyn til kostnader knyttet til produksjon av den minste pakken med tjenester og ikke kostnader som oppstår som følge av produksjon av tjenester i serviceanlegg, tilleggstjenester og ekstratjenester. Det er kun de trafikkavhengige kostnadene (direkte kostnader) som skal inngå grunnprisen.

I forordning 2015/909 artikkel 3 (4) nevnes det eksempler på hvilke typer kostnader som kan inkluderes i de direkte kostnadene, dersom infrastrukturforvalter «*på en åpen, utfyllende og objektiv måte, blant annet på grunnlag av beste internasjonale praksis, kan fastsette kostnader og godtgjøre at de er oppstått som en direkte følge av driften av togtjenesten*». Bane NOR har lagt til grunn kun vedlikeholdskostnader i de direkte kostnadene.

4.1.2 Valg av beregningsmetode

Neste trinn er valg av overordnet beregningsmetode. I artikkel 6 i forordning 2015/909 åpnes det for å beregne de direkte trafikkavhengige kostnadene ved minstepakken ved hjelp av følgende metoder:

- Aktivitetsbasert kostnadskalkyle (forordningsmetoden): Totale kostnader ved å tilby minstepakken, fratrukket ikke-kvalifiserte kostnader.
- Teknisk kostnadsmodellering (ingeniørmetoden)
- Økonometrisk metode

Ved bruk av den førstnevnte metoden, som vi har kalt forordningsmetoden, trekker man fra ikke-kvalifiserte kostnader i henhold til artikkel 4 i forordningen. I mange tilfeller vil dette innebære stor grad av skjønn når det gjelder å avgjøre i hvilken grad de ulike typene kostnader er trafikkavhengige eller ikke. Skjønnsbaserte vurderinger kan være utfordrende å dokumentere på en etterprøvbar måte.

Ingeniørmetoden kan for det første bestå av modeller for fysiske sammenhenger, for eksempel mellom trafikk med ulike togtyper og slitasje på ulike kategorier av infrastruktur. For det andre kan metoden innebære ulike måter å knytte kostnader til slitasjen – for eksempel ved å multiplisere slitasjemål med enhetskostnader, eller allokere kostnader fra regnskapsdata ut på de identifiserte driverne. På denne måten finner man estimater for trafikkavhengige kostnader per enhet, for eksempel vedlikeholdskostnader per togkilometer. I prosessen er man avhengig av både gode data og ekspertvurderinger.

Med økonometrisk metode baseres estimeringen av trafikkavhengige kostnader i stor grad på statistikk og blir enklere å etterprøve kvantitativt. Men det kreves også jernbanefaglige vurderinger for å spesifisere en god økonometrisk modell for hvordan kostnadene varierer med ulike forklaringsvariabler. Det er også viktig med god kvalitet på datasettet.

Det er styrker og svakheter ved både ingeniørmetoden og økonometrisk modellering. Smith et al. (2017) beskriver en tottrinnsmetode hvor teknisk og økonomisk modellering kombineres. Metoden er svært kompleks med hensyn til data og modellering. Bane NORs divisjon for Drift og teknologi arbeider med å være stadig mer datadrevet slik at for eksempel vedlikehold skal være basert på data om infrastrukturens tilstand på et avansert nivå. Det ville likevel være krevende å etablere en modell i tråd med opplegget fra Smith et al. Det er en mulighet som kan bli mer realistisk i framtidige revisjoner av infrastrukturavgiftene.

I denne analysen har vi valgt å gå videre med økonometrisk metode. Etter vår vurdering er det den metoden som i minst grad baseres på skjønn. Datasettet alene kan likevel ikke være det eneste man baserer seg på i etableringen av en økonometrisk modell. Modellutformingen har derfor skjedd i dialog mellom økonomer og jernbanefaglig kompetanse³. Professor Nils Henrik Von der Fehr ved Universitet i Oslo har kommet med nyttige innspill om regresjonsanalysen og hensyn som må tas i beregningen av ny grunnpris. I arbeidet med beregning av grunnpris i 2021 fikk Bane NOR flere tilbakemeldinger fra Statens jernbanetilsyn og professor Christian Riis ved BI, og det er forsøkt å hensynta også disse innspillene. Joanna Kiepiela som jobber med asset management i Bane NOR og har god kunnskap om forholdet mellom kostnader og vedlikehold i jernbanen har bidratt i mer prinsipielle jernbanefaglige spørsmål.

4.1.3 Veien videre

Valget av økonometrisk metode fører oss til de neste trinnene i fremgangsmåten i Figur 1.

- Valg av funksjonsform (kapittel 4.3.2). Eksempler på funksjonsformer i økonometriske analyser er dobbel-log, translog og kvadratisk form. Dette er ulike former for å modellere sammenhengen mellom på den ene siden kostnader og på den andre siden ulike forklaringsvariabler.
- Valg av variabler (kapittel 4.3.3). I tillegg til togtrafikken vil det være andre variabler som bidrar til å forklare variasjon i kostnader. Eksempler på mulige kandidater er tekniske variabler som tunneler, bruer, sporveksler, plattformer, kurvatur og hastighet. Andre typer

³ Arbeidet som er gjort og resultatene som fremgår er helt og holdent Bane NOR sitt ansvar

forklaringsvariabler kan være regioner (gruppering av banestrekninger), vektklasser eller togslog (gods- og persontog).

- Mål på trafikken, det vil si valg av enhet (kapittel 4.3.6). Kostnadene kan for eksempel uttrykkes som kroner per togkilometer, bruttotonnkilometer eller antall tog.
- Tester (Vedlegg D – tester av modellen) for å vurdere kvaliteten på resultatene av den økonometriske analysen.

4.2 Data

Beregningene bygger på et paneldatasett bestående av kostnadsdata, trafikkdata og tekniske data for 26 banestrekninger over de 5 årene 2017-2021⁴.

4.2.1 Kostnadsdata

Det er forebyggende og korrektivt vedlikehold som inngår i vedlikeholdskostnadene. Dataene er hentet fra Bane NORs regnskapssystem.

I forordningen er det under artikkel 4 listet opp kostnadselementer som ikke er kvalifiserte som direkte kostnader, og som dermed ikke skal være med i grunnprisen. Noe av dette er hensyntatt i uttaket av kostnadsdata, ved at f.eks. finanskostnader og overhead-kostnader ikke er inkludert. Ekskluderingen av andre kostnader, som f.eks. pkt. (1), bokstav o) «den del av kostnadene til vedlikehold og fornyelse av sivil infrastruktur som ikke oppstår som en direkte følge av driften av togtjenesten» overlates til regresjonsmodellen.

4.2.2 Tekniske data

Ved bruk av tekniske data i analysen får man tatt hensyn til hvordan infrastrukturens utforming påvirker marginalkostnadene. De tekniske dataene er hentet fra TRASE⁵. Dataene er fordelt på banestrekninger.

4.2.3 Trafikkdata

Produksjonsvolumet er målt gjennom faktisk kjørte brutto tonnkilometer, togkilometer og gjennomsnittlig vekt registrert i TIOS⁶ og DRAGE⁷ totalt og for henholdsvis persontog, godstog og andre tog. Vi har også data på gjennomsnittlig faktisk aksellast i 2020. Alle variablene er fordelt på banestrekning.

4.2.4 Prisindeks

Siden datasettet består av kostnader i løpende priser for forskjellige år, må det justeres for ulikt prisnivå. Dette tas hensyn til gjennom å enten inkludere en prisindeks som kontrollvariabel eller justere vedlikeholdskostnadene med prisindeksen direkte for å få kostnadene i faste priser. Det er ikke rett frem hva man skal bruke som mål på prisveksten. Det vanlige i Europa når det kommer til årlig prisjustering av infrastrukturavgiftene er å enten bruke konsumprisindeksen eller kostnadsindekser for jernbane eller lignende tjenester. I Bane NORs avtaler med Spordrift⁸ justeres prisene med SSBs

⁴ Datakvaliteten er bedre i disse årene enn de har vært tidligere. Helst skulle vi hatt en lengre tidsserie, og ved neste revisjon av prisene vil analysen kunne inkludere flere år.

⁵ Gir informasjon om jernbaneinfrastrukturen

⁶ Bane NORs Trafikkinformasjons- og oppfølgingssystem

⁷ DRAGE = Data, Rapport og Analyse Generator og er Bane NORs datavarehus for oppfølging av punktlighet, økonomi og beregning av infrastrukturavgifter

⁸ Totalleverandør innen drift og vedlikehold av banerelatert infrastruktur i Norge

prisindeks for drift og vedlikehold av veganlegg, og denne indeksen er dermed et godt mål på faktisk prisvekst i vedlikeholdskostnadene knyttet til jernbane i Norge.

4.2.5 Datasett

Kostnadsdata, trafikkdata og tekniske data på strekningsnivå, samt prisindeks ble samlet i et datasett. Vedlikeholdskostnadene i 2021-kroner inngikk i modellen som avhengig variabel, mens følgende uavhengige variabler var tilgjengelige i datasettet, og det ble vurdert om de var relevante eller ikke å inkludere i modellen.

Tabell 1 Datasett

Variabel	Enhet
Kostnadsdata	
Vedlikeholdskostnader	I kroner
Tekniske data	
Banelengde	Km med spor
Bruer	Konstruksjonslengde
Sporveksler	Antall
Tuneller	Konstruksjonslengde
Plattformer	Konstruksjonslengde
Kurvatur	Gjennomsnittlig kurvatur < 300 meters radius
Elektrifisering	Dummy for elektrifisert eller ikke
Skiltet hastighet	Gjennomsnittlig tillatt km/t
Trafikkdata	
Antall avganger	
Brutto tonn	
Aksellast	Gjennomsnitt
Prisindeks og tidstrend	
SSBs prisindeks for drift og vedlikehold av veganlegg	
Tidstrend	

4.2.6 Svakheter ved dataene

På noen områder er datasettet bedre nå enn ved tidligere analyser. Blant annet har flere analyser tidligere brukt planlagt trafikkarbeid som mål på trafikken⁹, mens vi i dag har gode data for faktisk trafikkarbeid. Likevel er det fortsatt noen svakheter ved datasettet.

Antall observasjoner

Datasettet består av totalt 130 observasjoner, med 5 år og 26 banestrekninger. I tillegg utelates tre observasjoner i modellen siden det er år med ingen trafikk på enkelte banestrekninger. Det at det ikke er flere datapunkter begrenser mulighetene en har i den økonometriske modellen. Med få observasjoner må man begrense hvor mange variabler som brukes i modellen. Jo flere variabler en har desto mindre variasjon blir det igjen i modellen.

Kostnadsdata

En svakhet ved datasettet er at ikke alle vedlikeholdskostnadene er fordelt på banestrekning. Noen kostnader er fordelt på banesjefs-nivå, mens en liten andel ikke er fordelt i det hele tatt. I disse tilfellene er kostnadene fordelt proporsjonalt med antall kilometer sporlengde.

Det har både vært en omlegging av økonomisystemet og en stor omorganisering i analyseperioden, noe som har medført koblingsutfordringer. Regnskapsmedarbeidere har bistått i arbeidet for å få så konsistente data over tid som mulig.

Regnskapsdataene er ikke laget med formålet om å beregne priser. Det er igangsatt et prosjekt i Bane NOR for å bl.a. få en bedre fordeling av kostnader fremover. Dette vil forhåpentligvis gjøre at vi får forbedrede data ved neste revisjon av prisene.

Kobling av data

For å få et samlet datasett er det benyttet flere datakilder, noe som har skapt noen utfordringer. I hovedsak gjelder dette den geografiske inndelingen, som er ulik for trafikkdata og økonomiske data. Dette har ført til at vi har endt opp med færre observasjoner enn hva som kunne vært mulig ellers. Det er etablert et masterdatateam i Bane NOR for å få en bedre struktur og kvalitet på masterdata. Dette vil antagelig gjøre slike koblinger av data enklere fremover.

Kollinearitet

Det kan være problemer med kollinearitet mellom forklaringsvariablene. Når en undersøker korrelasjon mellom variablene som potensielt kan inkluderes i modellen er det tydelig at dette er en utfordring. Dette kommer tydelig frem i korrelasjonsmatrisen i tabell 2, der det i noen tilfeller er en korrelasjon på opp mot 0,9. Sporlengde er en viktig forklaringsvariabel for kostnader, og denne er derfor viktig å inkludere i modellen. Både antall sporveksler, plattformlengde, meter med bru og meter med tunnel er høyt korrelert med sporlengde. En måte å omgå dette problemet på er å inkludere variablene som andeler av sporkilometer. Det ble kjørt modeller med denne spesifiseringen, men resultatet ble koeffisienter som var ikke-signifikante og hadde motsatt fortegn enn hva som gir mening jernbanefaglig. Dette taler imot å inkludere disse variablene.

⁹ Jernbaneverket 2014

Bane NOR Metodikk og beregninger av direkte kostnader

Tabell 2 Korrelasjon mellom variabler

	km	Tunell	Bru	Antall tog	sporveksel	plattform	kurvatur
Km	1						
Tunell	0,7442	1					
Bru	0,7368	0,6134	1				
Antall tog	-0,0842	0,0668	0,2826	1			
Sporveksel	0,8921	0,6703	0,8666	0,1251	1		
Plattform	0,8434	0,6787	0,8562	0,1691	0,9645	1	
Kurvatur	0,6125	0,5367	0,3231	-0,1492	0,5348	0,476	1

Tabell 3 Korrelasjon mellom variabler, med andeler

	Km	Tunnelandel	Bruandel	Antall tog	sporveksel	Plattformandel	kurvatur
Km	1						
Tunnelandel	0,0241	1					
Bruandel	-0,206	0,1201	1				
Antall tog	-0,0842	0,3586	0,4854	1			
Sporveksel	0,8921	0,0751	0,0209	0,1251	1		
Plattformandel	-0,0938	0,2355	0,3487	0,6206	0,1418	1	
Kurvatur	0,6125	0,1698	-0,2854	-0,1492	0,5348	-0,1203	1

4.3 Modell og vurderinger

Målet med den økonometriske metoden er å måle vedlikeholdskostnadene som en funksjon av trafikken, der det kontrolleres for diverse karakteristikker ved jernbanenettet. I estimeringen av direkte kostnader har Bane NOR blant annet testet ulike funksjonsformer og forskjellige mål på trafikkbelastning for å finne den modellen som best forklarer variasjonen i vedlikeholdskostnadene. Beregningene er gjort med bruk av statistikkpakken STATA.



Figur 2 Sammenheng mellom antall togkilometer og vedlikeholdskostnader

Modellen bygger på en forutsetning om at sammenhengen illustrert i Figur 2 er sann. Blant annet har eksperter i Bane NOR på sammenheng mellom vedlikehold og kostnader bekreftet både at antall tog

Metodikk og beregninger av direkte kostnader

påvirker slitasjen på infrastrukturen og at slitasjen på infrastrukturen er en viktig del av det som bestemmer vedlikeholdet og dermed vedlikeholdskostnadene. I tillegg kommer det frem i resultatene av den statistiske analysen at det også er signifikant statistisk sammenheng.

IRG-Rail¹⁰ anbefaler i en rapport fra 2016 følgende fremgangsmåte for å estimere direkte kostnader ved bruk av en økonometrisk metode:

Step	Aim	Description
1 a	Collect data on costs	<ul style="list-style-type: none"> - Cost data on operation, maintenance and renewal costs - A preferred approach is to disaggregate these costs by type - Should be collected at the network section level, for at least one year - Identifiable fixed and non-eligible costs may be excluded from the cost base
1 b	Collect data on traffic	<ul style="list-style-type: none"> - Should be collected at the same observation unit as costs, for the same period - Usually expressed in tonne-km or train-km and a preferred approach is to collect data on each traffic type
1 c	Collect other data	<ul style="list-style-type: none"> - Collect data on infrastructure characteristics (e.g. number of tracks, type of rail, age of different components) at the same level as cost and traffic data - Additional data such as the different climates or regions can also be collected
2	Assemble datasets	<ul style="list-style-type: none"> - Data on costs, traffic and infrastructure characteristics should be merged into a single dataset to enable econometric estimations
3	Econometric estimation	<ul style="list-style-type: none"> - Estimate econometric models where the dependent variables are the types of costs, the variables of interest are the traffic types and control variables include for instance infrastructure characteristics, climate and region variables
4a	Derive elasticity	<ul style="list-style-type: none"> - From the results of the econometric model, it is possible to estimate the elasticity of costs to traffic - The elasticity represents the percentage of costs that vary with the level of traffic.
4b	Calculate marginal costs	<ul style="list-style-type: none"> - Marginal costs should then be calculated. They represent the costs per unit of traffic considered (e.g., per train-km or (gross)tonne-km) - If an elasticity has already been derived then the marginal costs may be calculated as elasticities times average costs

Figur 3 Estimering av direkte kostnader ved bruk av en økonometrisk metode

Det er denne fremgangsmåten Bane NOR har tatt utgangspunkt i ved beregningen av direkte kostnader.

4.3.1 Forutsetninger for regresjonsmodellen

I estimeringen av regresjonsmodellen brukes minste kvadraters metode. Denne metoden går ut på å minimere «avstanden» mellom observasjonspunktene og den estimerte linja. Man minimerer altså residualleddene gitt ved:

$$Q = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (Y_i - a - bx_i)^2$$

Der Q er det vi ønsker å minimere ved å estimere a og b, e_i er residualledd for observasjon i, $Y_i = a + bx_i$ er regresjonslinja der Y_i er avhengig variabel, a konstant, b stigningstallet og x_i er uavhengig variabel. I

¹⁰ IRG–Rail Charges Working Group, An introduction to the calculation of direct costs in respect of implementing regulation 2015/909 9 November 2016

ligningen over er det kun en forklaringsvariabel, men det samme prinsippet gjelder i en multipl regresjonsmodell.

Det er noen forutsetninger som bør være oppfylt for å kunne bruke en regresjonsmodell med minste kvadraters metode:

- Lineær sammenheng mellom variablene
- Residualleddene skal ha konstant varians (homoskedastisitet)
- Residualleddene skal være normalfordelte
- Residualleddene skal være uavhengige (ingen autokorrelasjon)

I Vedlegg D – tester av modellen undersøkes det om disse forutsetningene er oppfylt.

4.3.2 Vurdering av funksjonsform

De mest vanlige tilnæringsmetodene i beregningen av marginalkostnader for jernbanen er bruk av translog eller dobbel-log funksjoner¹¹. I realiteten innebærer disse funksjonsformene at man estimerer enten marginalkostnadene eller kostnadselastisiteten uavhengig av de tekniske og fysiske sammenhengene.

Translog

Det optimale hadde vært å velge en funksjonsform som har minimalt med restriksjoner pålagt på den underliggende teknologien, f.eks. translog. En slik funksjonsform krever imidlertid at flere parametere skal estimeres, noe som gjør at det blir færre frihetsgrader igjen¹².

Dobbel-log

En dobbel-log-funksjonsform er et alternativ til translog. Ulempen er at den er mer restriktiv siden det er en antagelse om konstant elastisitet, men fordelene er bl.a. at resultatene av modellen er enkle å tolke.

Kvadratisk modell

En kvadratisk modell kan inkludere forklaringsvariablene enkeltvis og multiplisert med seg selv, for eksempel formen $Z = b_0 + b_1X + b_2Y + b_3X^2 + b_4Y^2 + \varepsilon$. Denne funksjonsformen tar høyde for ikke-lineære sammenhenger.

Resultater

Funksjonsformene dobbel-log, translog og kvadratisk modell ble testet, herunder ulike varianter av hver av dem med hensyn til hva slags trafikkmål og andre variabler som skulle inngå i modellen. I tabell 6 i Vedlegg B – Regresjonsresultater ulike modeller vises resultater fra et utvalg av testene.

I valget mellom dobbel-log, translog og kvadratisk modell ble flere kriterier vurdert. Justert R^2 er et mål på hvor mye av variasjonen i kostnader som kan forklares av variablene i modellen. Justert R^2 er ikke direkte sammenlignbar på tvers av de ulike funksjonsformene. Alle har et akseptabelt nivå, og det er for små forskjeller til at R^2 kan vektlegges i noen særlig grad i valget av modell. Videre så vi på om modellen ga signifikante koeffisienter. Her viste testene for translog seg lite egnet, da viktige variabler som kilometer og trafikkmål (antall tog eller bruttotonn) ikke blir signifikante. Translog så følgelig ikke ut til å være en god modell for å beskrive sammenhengene mellom togtrafikken og kostnadene i vårt

¹¹ Se f.eks. tabell 14 i Catrin 2008.

¹² Catrin 2008

datasett¹³, og ble derfor ikke valgt som funksjonsform. Kvadratisk modell ga signifikante koeffisienter for de nevnte variablene, men det er problematisk å gi resultatet for marginalkostnader en meningsfylt tolkning i de variantene vi har testet. I tillegg ble enkelte koeffisienter ikke-signifikante.

4.3.2.1 Dobbel-log modellen

Utfra vurderingene over falt valget på en dobbel-log-modell. I en slik modell er det elasticiteten man estimerer.

Rent matematisk vil det være en sammenheng mellom marginalkostnaden og gjennomsnittskostnaden via kostnadselasticiteten. Sammenhengen er:

$$B' = e(x) * \bar{B}(X)$$

B' = marginalkostnaden

\bar{B} = gjennomsnittskostnaden

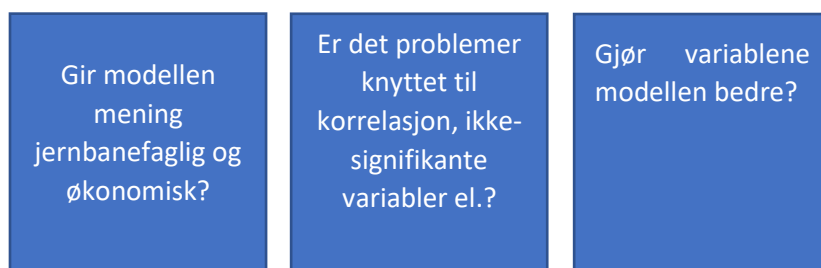
e = kostnadselasticiteten

x = produksjonsvolum

Kostnadselasticiteten angir i hvilken grad totalkostnadene øker med økende produksjonsvolum. En prosent økning i produksjonsvolum gir $e(x)$ prosent økning i totalkostnadene. Dersom elasticiteten er mindre enn én ($e(x) < 1$), vil vi ha fallende gjennomsnittskostnader med hensyn til produksjonsvolum, eller stordriftsfordeler. Prising basert på marginalkostnadsprinsippet vil i dette tilfelle ikke gi full kostnadsdekning. Dersom elasticiteten er større enn én ($e(x) > 1$), vil marginalkostnaden være større enn gjennomsnittskostnaden, og prising basert på marginalkostnadsprinsippet vil gi kostnadsdekning. Jernbanen er kjennetegnet ved fallende gjennomsnittskostnader¹⁴, så priser basert på marginalkostnader vil ikke gi kostnadsdekning.

4.3.3 Vurdering av hvilke variabler som skal inngå i modellen

I figur 3 vises noen av vurderingene Bane NOR har tatt i spesifiseringen av modellen.



Figur 4 Vurdering av variabler

Hvilke variabler kan være relevante å inkludere?

Tekniske variabler

Jernbanefaglig kan det være relevant å inkludere alle de tekniske variablene som er listet opp i tabell 1.

¹³ Det samme var f.eks. tilfellet for Daljord 2003

¹⁴ Jernbanedirektoratet 2018

Som det fremkommer i kapittel 4.2.6 er det utfordringer knyttet til å inkludere de tekniske variablene. Det er høy korrelasjon mellom variablene, og dette taler imot å inkludere flere av de tekniske variablene.

Tidstrend

Det ble testet modeller der en trendvariabel/tidsindeks ble inkludert i modellen¹⁵. Trendvariabelen kan brukes som en proxy på en variabel som påvirker den avhengige variabelen og som ikke er direkte observerbar, men som er høyt korrelert med tiden. Det å inkludere en tidstrend gjorde nokså små utslag på de andre koeffisientene, og en sammenligning av modellene i STATA viste at det å inkludere en tidstrend gjorde modellen dårligere. Se resultat av sammenligningen i vedlegg C.

Tidsforskyvning

Det ble vurdert om en burde ha med en tidsforskyvning i datasettet. Dersom kostnadene i år t i stor grad avhenger av kostnadene/trafikken i år $t-1$, vil dette være lurt å kontrollere for. Ekspertene i Bane NOR mener imidlertid at dette ikke er tilfellet. I tillegg vil man miste et helt år med observasjoner dersom man velger å ta det med. Derfor falt valget på å ikke inkludere noen tidsforskyvning i modellen.

4.3.4 Differensiering mellom baneområder/regioner

Ved å dele banestrekningene inn i regioner, kan man både hensynta at det er ulike kostnadsnivåer og man får indirekte justert for at det forskjeller i de tekniske variablene uten å ta med disse direkte i modellen. I tillegg kan man få med ikke-observerbare forskjeller mellom regionene.

Det ble undersøkt om det var signifikant forskjellige nivåer på vedlikeholdskostnadene eller ulik sammenheng mellom trafikk og vedlikeholdskostnader i forskjellige regioner ved å dele banestrekningene inn i tre regioner.

Valg av inndeling

Det ble undersøkt om det er noen karakteristikk ved banene som gjorde at det var en naturlig inndeling i baneområder. Etter en samlet vurdering av de tekniske variablene, kostnadsnivået og mengden trafikk, landet vi på følgende inndeling:

- Osloområdet
- Oftobanen
- Andre baner

I tidligere studier har nettet vært delt inn i flere regioner. Utfra en vurdering av de tekniske variablene mv. finner vi imidlertid at det ikke er grunnlag for å dele nettet opp ytterligere. I tillegg er det vesentlig at modellen viser at det er signifikante forskjeller i kostnader mellom regionene for at det skal rettferdiggjøres å differensiere prisene, jf. artikkel 5 (1) i forordningen. Dette er f.eks. ikke tilfelle i en tidligere analyse av Jernbaneverket¹⁶.

Heterogenitet i konstanten eller stigningskoeffisienten

Det er to måter å sjekke om ulike baneområder har ulik elasticitet, enten ved å inkludere dummyvariabler for baneområdene eller ved å inkludere baneområdene som interaksjonsvariabler. Å inkludere dummyvariabler tilsvarer å skifte konstanten i regresjonslikningen, dvs. at regresjonslinja får ulikt nivå for de ulike baneområdene. Å inkludere baneområdene som interaksjonsvariabler vil si at

¹⁵ <http://www.econometrics.com/intro/trend.htm>

¹⁶ Kjørevegsavgift for norsk jernbane - Metodikk og resultater 2014

variablene som representerer baneområder multipliseres med trafikkvariabelen. Da er det koeffisienten for trafikk som blir forskjellig for hvert baneområde. Vi ønsket å analysere om det er forskjell mellom baneområdene når det gjelder i hvilken grad en endring i trafikk gir utslag i endring i kostnader. Interaksjonsvariabler er en mer treffsikker metode til dette formålet.

4.3.5 Justering av prisnivå

Når prisindeksen ble inkludert som en egen variabel fikk denne en veldig høy elasticitet. Dette kan tyde på at det er noe annet enn kun prisnivået som fanges opp i denne variabelen. På grunn av dette ble vedlikeholdskostnader i faste priser inkludert i modellen i stedet. Kostnadene i faste priser ble beregnet ved å justere kostnadene i løpende priser med SSBs prisindeks for drift og vedlikehold av veganlegg. Denne endringen førte ikke til noen vesentlig endring i resultatene.

4.3.6 Mål på trafikken

I artikkel 5 (1) i forordningen står det at «Infrastrukturforvalteren skal beregne gjennomsnittlige enhetskostnader for hele jernbanenettet ved å dividere de direkte kostnadene for hele nettet med det samlede antallet forventede eller faktisk gjennomførte kjøretøykilometer, togkilometer eller bruttotonnkilometer». Bane NOR har valgt å bruke togkilometer som mål på trafikken. Det ble kjørt ulike modeller med togkilometer og brutto tonnkilometer som mål på trafikkbelastningen i regresjonsmodellen, men dette ga ikke signifikant forskjell i modellens forklaringsmakt. Togkilometer er også brukt som prisenhet i mange andre land i Europa¹⁷. Det ble vurdert om en skulle ha ulike priser for ulike vektclasser, men siden prisene differensieres mellom regioner, var ikke dette mulig i datasettet da korrelasjonen mellom region og vektklasse er svært høy.

4.3.7 Valgt modell

Da valget av funksjonsform falt på en log-log-variant og togkilometer ble valgt som estimat på trafikkbelastning, ble modellen for estimering av elasticiteter følgende:

$$\ln \text{Fastkost} = \beta_0 + \beta_1 \ln Km + \beta_2 \ln \text{Antalltog} + \beta_3 \text{Osloområdet} * \ln \text{Antalltog} + \beta_4 \text{Oforbanen} * \ln \text{Antalltog} + \epsilon$$

Osloområdet og *Oforbanen* kan ha verdien 0 eller 1. For andre banestrekninger er begge lik 0, og koeffisienten for $\ln \text{Antalltog}$ er β_2 . For banestrekninger i Osloområdet blir koeffisienten til $\ln \text{Antalltog}$ lik $\beta_2 + \beta_3$, mens for Oforbanen blir koeffisienten $\beta_2 + \beta_4$.

I Tabell 4 presenteres deskriptiv statistikk over variablene som ble vurdert i modellen (se Tabell 1 for måleenheter).

¹⁷ IRG-rail 2020, Review of charging practices for the minimum access package in Europe

Bane NOR
Metodikk og beregninger av direkte kostnader

Tabell 4 Deskriptiv statistikk¹⁸

Variabel	Antall observasjoner	Gjennomsnitt	Standardavvik	Min	Max
Fastkost	130	78 700 000	80 500 000	762 823	293 000 000
Km	130	189	228	2	801
Antall tog	130	14 534	21 834	-	126 856
Brutto tonnkm	130	5 345 866	6 824 016	-	34 100 000
Sporveksel	130	103	119	-	470
Platt	130	5 300	6 367	-	27 434
Tunnel	130	13 139	22 478	-	78 803
Bru	130	3 330	3 761	71	12 671
Kurvatur	130	94	101	1	398
Gjennomsnittlig tillatt snitthastighet	130	79	21	25	136
Gjennomsnittlig aksellast	130	14	2	11	18

4.3.8 Resultater av regresjonsmodellen

I tabell 6 presenteres estimeringsresultatene basert på vanlig minste kvadraters metode med den valgte dobbel-log funksjonsformen. Se vedlegg B for resultater med andre modellvalg og vedlegg D for diverse tester av resultatene.

Tabell 5 Regresjonsresultater med valgt modell

In_fastkost	Koeffisient	Standardavvik	T-verdi	P> t	Konf.intervall Min	Konf.intervall maks
Log(antall tog)	0,1569292	0,0603735	2,6	0,01	0,0374138	0,2764446
Log(km)	0,9196963	0,0403052	22,82	0	0,8399081	0,9994845
Interaksjon Ofotbanen	0,1212635	0,0311231	3,9	0	0,0596523	0,1828747
Interaksjon Osloområdet	0,0484541	0,0180611	2,68	0,008	0,0127004	0,0842077
Konstant	11,78138	0,4776142	24,67	0	10,83589	12,72686

Marginalkostnadene utledes utfra koeffisienten for antall avganger og regionsvariablene. Elastisiteten for antall tog på 0,157 gjelder for referansegruppen, som er region «andre strekninger». Tolkningen av

¹⁸ Elastisiteten for antall togkm vil være lik den for antall tog, da banelengde er konstant fra år til år

en elastisitet på 0,157 for antall tog er at en 10 % økning i antall tog/togkilometer gir 15,7 % i økte kostnader. For togene som kjører i region 2, Ofotbanen, vil elastisiteten for togkm være 0,278¹⁹. Tilsvarende vil elastisiteten for togkm være 0,205 i region 3, Osloområdet. Kombinert med trafikkdata blir den vektete elastisiteten 0,173. Resultatet stemmer godt overens med tidligere studier²⁰. Utfra justert R² beskriver modellen 84,1 % av variasjonen i kostnadene. Elastisitetene kombinert med gjennomsnittskostnaden gir følgende gjennomsnittlige marginalkostnader for hver region (se avsnitt [4.3.1.1](#) for sammenheng mellom elastisitet og gjennomsnittskostnader):

Tabell 6 Vedlikeholdskostnader per togkm for ulike regioner, i kroner

	Elastisitet	Togkm	Kostnader	Gjennomsnittskostnad	Marginalkostnad, kroner per togkm
Osloområdet	0,21	14 319 646	326 476 096	22,8	4,7
Ofotbanen	0,28	353 556	59 607 699	168,6	46,9
Resten	0,16	32 130 681	1 639 772 660	51,0	8,0

5 Resultater

Prisen vil bli bestemt utfra følgende ligning:

$$P = p * togkm$$

Hvor P er pris per avgang, p er pris per togkilometer og togkm er faktisk kjørte togkilometer.

Med p gitt ved følgende tabell:

Tabell 7 . Kroner per togkilometer for hver region

Region	Marginalkostnad
Osloområdet	4,7
Ofotbanen	46,9
Andre strekninger	8,0

6 Implementering og anbefalinger

Denne rapporten vil inngå i høringen av Network Statement 2024 og det tas sikte på at ny prismodell og resultatene i denne rapporten vil bli gjeldende fra og med ruteplanperiode 2024.

Bane NOR legger ikke opp til at det blir noen implementeringsperiode.

Bane NOR legger i utgangspunktet opp til at prisene justeres periodisk. Ved vesentlig bedre grunnlagsdata eller andre større endringer vil prisene imidlertid kunne endres på bakgrunn av dette.

¹⁹ Elastisitet referansegruppe pluss elastisitet for gruppen.

²⁰ Se f.eks. CATRIN 2008

Periode mm	Beskrivelse
Femårig justering	<p>Bane NOR vil oppdatere kostnadskalkylene om lag hvert femte år basert på tilsvarende eller forbedrede metoder, og der mer oppdaterte data kan brukes som grunnlag i estimeringen. Mellom de større justeringene endres prisene årlig i henhold til en egnet SSB-indeks. Det benyttes kostnadsindeksen for drift og vedlikehold av veganlegg.</p> <p>Selve prisjusteringen foretas etter følgende prinsipp (1):</p> $(1) \quad P_{t+1} = P_t \cdot \left(\frac{KI_t^{Q2}}{KI_{t-1}^{Q2}} \right)$
Årlig justering	<p>der: P_{t+1} = pris neste år P_t = pris inneværende år KI^{Q2} = SSBs indeks pr. annet kvartal for inneværende (t) og foregående (t-1) år</p> <p>Dette innebærer en prisjustering etterskuddsvis, men den gir stor forutsigbarhet for togselskapene, da neste års priser vil være klare tredje kvartal året før. Samtidig kan man følge med på indeksen underveis i året.</p>
Nye, ombygde eller nedlagte objekter	<p>Dersom det i forbindelse med nye anlegg, ferdigstilles nye objekter eller større ombygninger av objekter, samt nedleggelse av gamle i fireårsperioden, skal dette tas inn i kostnadsgrunnlaget når anlegget/objektet tas i bruk eller tas ut av bruk.</p>

7 Bibliografi

Infrastrukturavgifter - Implementeringsplan, Gunnar Markussen 2017

IRG-Rail, Review of charging practices for the minimum access package in Europe, November 2020

CATRIN Cost Allocation of TRansport INfrastructure cost, Mars 2008

Track Access Charges in Freight Transport, Justina Hudenko, Cambridge Scholars Publishing 2019

<https://stats.idre.ucla.edu/other/mult-pkg/faq/general/faqhow-do-i-interpret-a-regression-model-when-some-variables-are-log-transformed/>

Kjørevegsavgift for norsk jernbane - Metodikk og resultater, Jernbaneverket, 2014

Marginalkostnader i jernbanenettet, Frichsenteret, Øystein Børnes Daljord, 2003

COMMISSION IMPLEMENTING REGULATION (EU) 2015/909 of 12 June 2015

Independent Regulators' Group – Rail IRG–Rail Charges Working Group, An introduction to the calculation of direct costs in respect of implementing regulation 2015/909 9 November 2016

<http://www.econometrics.com/intro/trend.htm>

Bane NOR

Metodikk og beregninger av direkte kostnader

Taboga, Marco (2021). "Heteroskedasticity-robust standard errors", Lectures on probability theory and mathematical statistics. Kindle Direct Publishing. Online appendix.

Jernbanedirektoratet, Veileder i samfunnsøkonomiske analyser i jernbanesektoren, 2018

8 Vedlegg A – Begreper og uttrykk

DRAGE = Data, Rapport og Analyse Generator og er Bane NORs datavarehus for oppfølging av punktlighet, økonomi og fakturering av infrastrukturavgifter (System for beregning av påslag ut fra data mottatt fra TIOS og ulike regler (beløpssatser mv.))

TRASÉ = TRAfikkrelatert StrEkningsinformasjon. TRASÉ skal gi tilgang til informasjon om infrastrukturen som er relevant for trafikkformål.

BaneData = Bane NORs infrastrukturdatabase

TIOS = Bane NORs Trafikkinformasjons- og oppfølgingssystem

Den minste pakken med tjenester = tjenester som nevnt i jernbaneforskriften § 4-1

Direkte kostnader = trafikkavhengige kostnader, dvs. kostnaden som oppstår som en direkte følge av den enkelte togtjenesten

Direkte enhetskostnad = direkte kostnad per togkilometer, kjøretøykilometer, et togs bruttotonnkilometer eller en kombinasjon av disse

Grunnpris = pris for den minste pakken med tjenester, tilsvarer de direkte kostnadene

Ikke-kvalifiserte kostnader = kostnadselementer som ifølge artikkel 4 i forordning 2015/909 ikke skal inngå i de direkte kostnadene

Aktivitetsbasert kostnadskalkulasjon (ABC-metode) = en metode for å henføre en bedrifts kostnader. Først henføres de til aktivitetene, og deretter videre til produktene basert på hvor mye produktene fungerer som en kostnadsdriver for de ulike aktivitetene.

Forordningsmetode = en form for ABC-metode som beskrives i forordning 2015/909

Økonometrisk metode = anvendelse av statistiske metoder, særlig for analyse av økonomiske data for å få innsikt i økonomiske forhold

Teknisk kostnadsmodellering = en form for ingeniørm metode, der man bruker ekspertvurderinger til å estimere kostnadssammenhenger

IRG-Rail = "Independent Regulators' Group – Rail", nettverk for reguleringsorganer fra 31 land i Europa

9 Vedlegg B – Regresjonsresultater ulike modeller

Tabell 8 Regresjonsresultater fra utvalgte modeller. Koeffisienter og tilhørende p-verdi (signifikanskriterium $p < 0,05$)

	1) Dobbellog med brutto tonn og interaksjon region	2) Dobbellog med antall tog og interaksjon region	3) Dobbellog med antall tog interaksjon region og trend	4) Dobbellog med antall tog og andel tekniske variabler	5) Translog med antall tog	6) Translog med bruttotonn og regioner	7) Kvadratisk, mill. kr., med mill. togkm og regioner
Justert R2	0,853	0,841	0,847	0,826	0,829	0,876	0,783
In_km	0,919	0,920	0,919	0,935	0,493	1,774	
p-verdi	(0)	(0)	(0)	(0)	(0,277)	(0,002)	
In_anttog		0,157	0,168	0,240	-0,147		

Bane NOR

Metodikk og beregninger av direkte kostnader

	1) Dobbellog med brutto tonn og interaksjon region	2) Dobbellog med antall tog og interaksjon region	3) Dobbellog med antall tog interaksjon region og trend	4) Dobbellog med antall tog og andel tekniske variabler	5) Translog med antall tog	6) Translog med bruttotonn og regioner	7) Kvadratisk, mill. kr., med mill. togkm og regioner
<i>p-verdi</i>		(0,01)	(0,006)	(0,011)	(0,771)		
ln_brt	0,175					-0,674	
<i>p-verdi</i>	(0)					(0,273)	
halv_Inkm2					0,142	0,123	
<i>p-verdi</i>					(0,002)	(0,002)	
halv_Intog2					0,057		
<i>p-verdi</i>					(0,296)		
Inkm_Intog					-0,023		
<i>p-verdi</i>					(0,64)		
halv_Inbrt2						0,090	
<i>p-verdi</i>						(0,039)	
Inkm_Inbrt						-0,093	
<i>p-verdi</i>						(0,006)	
mill_togkm							71,041
<i>p-verdi</i>							(0)
mill_togkm2							-6,595
<i>p-verdi</i>							(0)
plattandel				-0,001			
<i>p-verdi</i>				(0,761)			
tunnandel				-0,001			
<i>p-verdi</i>				(0,45)			
Bruandel				0,006			
<i>p-verdi</i>				(0,056)			
ln_kurv1				-0,051			
<i>p-verdi</i>				(0,365)			
doslo_Intog		0,048	0,046				
<i>p-verdi</i>		(0,008)	(0,01)				
dofot_Intog		0,121	0,121				
<i>p-verdi</i>		(0)	(0)				
doslo_Inbrt	0,030						
<i>p-verdi</i>	(0,006)						
dofot_Inbrt	0,038						
<i>p-verdi</i>	(0,033)						
Osloområdet						0,226	-61,235
<i>p-verdi</i>						(0,273)	(0)
Ofofbanen						0,323	23,187
<i>p-verdi</i>						(0,338)	(0,183)
aar2			0,120	0,123			
<i>p-verdi</i>			(0,469)	(0,488)			
aar3			0,298	0,299			
<i>p-verdi</i>			(0,069)	(0,087)			
aar4			0,396	0,403			
<i>p-verdi</i>			(0,016)	(0,022)			
aar5			0,359	0,363			
<i>p-verdi</i>			(0,029)	(0,038)			
Konstant	10,63	11,78	11,45	10,99	13,57	14,29	12,76
<i>p-verdi</i>	(0)	(0)	(0)	(0)	(0)	(0,003)	(0,013)

10 Vedlegg C – modell med og uten tidstrend

Sammenligningen av modell med og uten tidstrend ble gjort i STATA med kommandoen nestreg. Som det kommer frem i tabellen under ga det en noe høyere justert R^2 å inkludere dummyvariabler for trend, men det totale signifikansnivået er høyest i modell 1. Enkelte av årene fikk også ikke-signifikante koeffisienter. Siden det i tillegg er liten forskjell i de viktige koeffisientene mellom modellene, ble modellen uten tidstrend valgt.

Tabell 9 Sammenligning av modell med og uten trend

Block 1: ln_anttog ln_km dofot_lntog doslo_lntog						
Source	SS	df	MS	Number of obs	=	127
Model	233.543119	4	58.3857797	F(4, 122)	=	167.93
Residual	42.4164059	122	.347675458	Prob > F	=	0.0000
				R-squared	=	0.8463
				Adj R-squared	=	0.8413
Total	275.959525	126	2.19015496	Root MSE	=	.58964

ln_fastkost	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
ln_anttog	.1569292	.0603735	2.60	0.010	.0374138	.2764446
ln_km	.9196963	.0403052	22.82	0.000	.8399081	.9994845
dofot_lntog	.1212635	.0311231	3.90	0.000	.0596523	.1828747
doslo_lntog	.0484541	.0180611	2.68	0.008	.0127004	.0842077
_cons	11.78138	.4776142	24.67	0.000	10.83589	12.72686

Block 2: aar2 aar3 aar4 aar5						
Source	SS	df	MS	Number of obs	=	127
Model	236.397465	8	29.5496831	F(8, 118)	=	88.14
Residual	39.5620601	118	.335271696	Prob > F	=	0.0000
				R-squared	=	0.8566
				Adj R-squared	=	0.8469
Total	275.959525	126	2.19015496	Root MSE	=	.57903

ln_fastkost	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
ln_anttog	.1677831	.0594236	2.82	0.006	.0501081	.285458
ln_km	.9194341	.0395818	23.23	0.000	.8410513	.9978169
dofot_lntog	.1212232	.0305631	3.97	0.000	.0606999	.1817465
doslo_lntog	.0461819	.0177542	2.60	0.010	.0110237	.0813401
aar2	.1203193	.1656421	0.73	0.469	-.2076971	.4483357
aar3	.2978713	.1623372	1.83	0.069	-.0236006	.6193431
aar4	.395842	.1625562	2.44	0.016	.0739365	.7177476
aar5	.3585304	.1624487	2.21	0.029	.0368377	.6802231
cons	11.45284	.4880582	23.47	0.000	10.48636	12.41933

Block	Block F	Block Residual df	Residual df	Pr > F	R2	Change in R2
1	167.93	4	122	0.0000	0.8463	
2	2.13	4	118	0.0815	0.8566	0.0103

11 Vedlegg D – tester av modellen

Bane NOR har gjennomført noen statistiske tester for å vurdere kvaliteten på modellen.

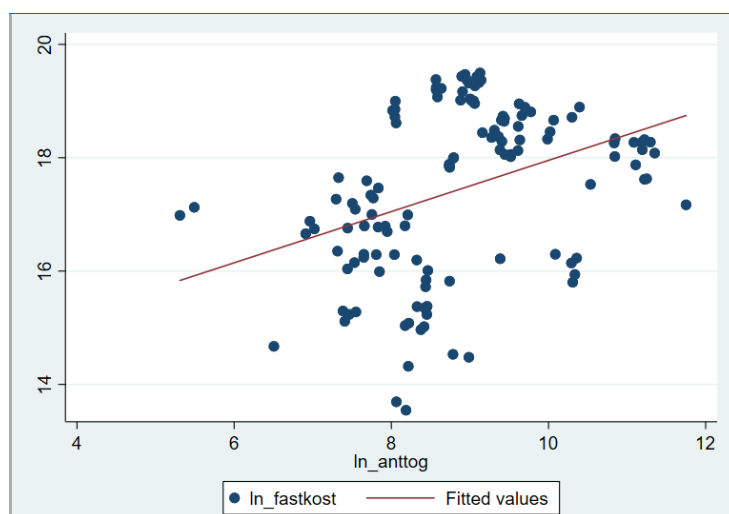
I en F-test ble nullhypotesen om at alle koeffisienter er lik 0 samtidig forkastet, og vi kan dermed konkludere med at korrelasjonen mellom modellen og de avhengige variablene er statistisk signifikant.

Det er også noen forutsetninger som bør være oppfylt for at en skal kunne bruke minste kvadraters metode:

- Lineær sammenheng mellom variablene
- Residualene skal ha konstant varians (homoskedastisitet)
- Residualene skal være normalfordelte
- Residualene skal være uavhengige (ingen autokorrelasjon)

Er sammenhengen mellom variablene lineær?

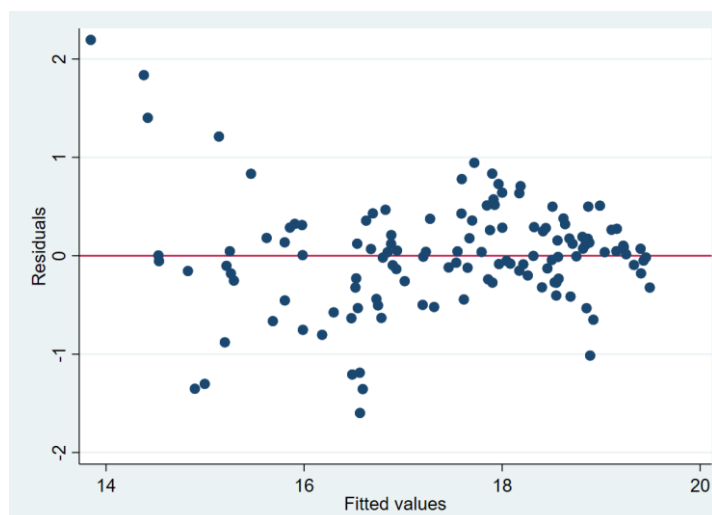
For å kunne bruke minste kvadraters metode må variablene i modellen ha en lineær sammenheng. I analysen er det gjort en antagelse om at det ikke er en lineær sammenheng mellom kostnader og antall tog, men når vi transformerer variablene til logaritmisk form blir likevel modellen lineær i de utledede variablene. For å teste om sammenhengen mellom de utledede variablene $\log(\text{faste kostnader})$ og $\log(\text{antall tog})$ er lineær er det tatt ut et plott fra STATA over disse to variablene. Figur 4. viser at en lineær modell er en god tilnærming.



Figur 5 Plott av $\log(\text{faste kostnader})$ og $\log(\text{antall tog})$

Har residualleddene konstant varians?

Det ble undersøkt hvorvidt antagelsen om konstant varians er oppfylt i valgt modell. For at denne antagelsen skal være oppfylt, skal plottet vise en jevn spredning omkring null, og spredningen skal ikke endre seg langs aksene. Dersom spredningen f.eks. bøyer seg ut i vifteform langs aksene (variansen blir større), bryter dette mot forutsetningene om konstant varians. Residualplottet viser at det er antydninger til noe større varians blant de lavere verdiene.



Figur 6 Residualplott med dobbel-log funksjonsform

Når variansen i residualleddene ikke er konstant, betyr det at man har utfordringer med heteroskedastisitet. Dette påvirker ikke estimatene eller justert R^2 , men det påvirker standardfeil, t-verdier og konfidensintervall. For å være sikker på at resultatene er signifikante også dersom det er et problem med heteroskedastiske restledd, ble det kjørt modeller i STATA med robuste standardavvik justert for heteroskedastisitet²¹. Koeffisientene var fremdeles signifikante. Her er regresjonsresultatene av den valgte modellen med robuste standardfeil for å justere for problemer med heteroskedastisitet.

```
Linear regression                               Number of obs   =       127
                                                F(4, 122)      =       110.40
                                                Prob > F       =       0.0000
                                                R-squared     =       0.8463
                                                Root MSE     =       .58964
```

ln_fastkost	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]	
ln_anttog	.1569292	.0587717	2.67	0.009	.0405846	.2732737
ln_km	.9196963	.0483248	19.03	0.000	.8240326	1.01536
dofot_lntog	.1212635	.0094018	12.90	0.000	.1026518	.1398752
doslo_lntog	.0484541	.019195	2.52	0.013	.0104556	.0864525
_cons	11.78138	.5710556	20.63	0.000	10.65092	12.91184

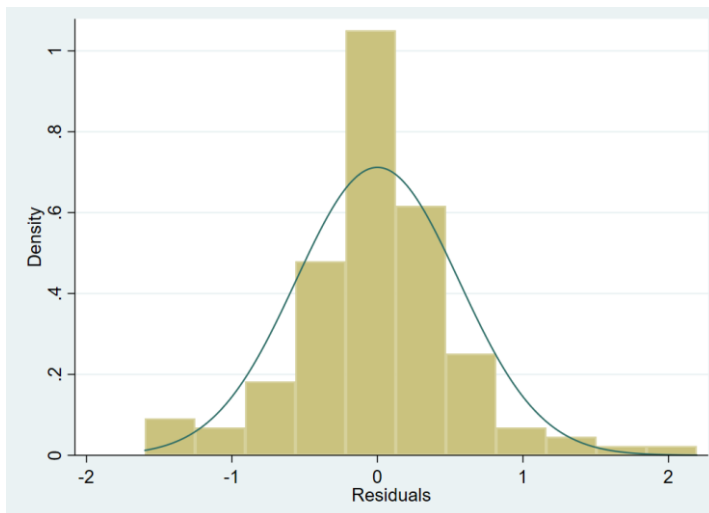
Figur 7 Regresjonsresultater med robuste standardfeil

Er residualleddene normalfordelte?

For å undersøke om residualleddene er normalfordelte, ble det laget et histogram i STATA. Dersom en skal kunne si at residualleddene er normalfordelte skal histogrammet ligne normalfordelingskurven, det vil si ha en symmetrisk topp sentrert i null. Figur 6 viser at det er normalfordeling i restleddene.

²¹ Taboga 2001

Bane NOR

Metodikk og beregninger av direkte kostnader

Figur 8 Fordeling av restledd

Er det problemer med autokorrelasjon?

En Durbin-Watson test indikerer at det ikke er autokorrelasjon i dataene, Durbin-Watson d-statistic på 2,19.