

Forsøg med redefinering af buskundepotentiale på Fyn

Thorbjørn Revsbech Sørensen, Diana Andreea Vasile, Patrycja Anna Zieba, Flemming Albæk og Seher Øzden.
FynBus

Baggrund

Siden 2015 har FynBus benyttet en model der beskriver potentielle kunder på Fyn. Modellen er sammensat af socioøkonomiske data, og forbrugeradfærdsdata som er definerede af flere registre og omdannet til et datasæt kaldet ConZoom™ (Geomatic, 2023). Modellen kaldes i FynBus regi for Buskundepotentiale (BKP).

I 2023 valgte Geomatic at opdatere grundlaget for ConZoom™. Der er derfor behov for en revidering af buskundepotentiale-modellen, især med henblik på de seneste opdateringer fra generation 5 til 6. Dette medfører at FynBus skal redefinere buskundeværdi (BKV) fra modellen fremadrettet med afsæt i ændringerne fra generation 5 til 6 af ConZoom™ og FynBus' passagerdata post Covid-19.

BKP kortlægger potentielle buskunder med udgangspunkt i ConZoom™, demografiske data og faglige vurderinger. BKP definerer det samlede potentiale af buskunder, såfremt de ville have adgang til den bedst mulige rejse til deres individuelle behov. Det er altså ikke en model som estimerer et passagertal, men estimerer potentiale ud fra bedste forudsætninger for individets rejsebehov.

Modellen finder brugbar anvendelse som arbejdsværktøj til at vurdere nye ruters teoretiske kundegrundlag i områder hvor der ikke tidligere har kørt kollektiv transport, og vurdering i forhold til rutens præstation i henhold til de kundesegmenter der bor og arbejder i områderne. Grundet udviklingen i data og datahøst på det kollektive område, er det muligt at anvende nye værktøjer til at forsøge at gøre buskundepotentialet bedre og dynamisk i forhold til strømninger i samfundet og sikre en bedre model fremadrettet.

Databeskrivelse

Conzoom data, skabes af Geomatic og er et overblik over persontyper og grupper, som kan benyttes ved analyse, der involverer adfærd, målgruppearbejde, analyser som involverer socioøkonomiske elementer osv. Der er i alt 9 grupper og 39 typer fordelt på disse grupper. Conzoom datasættet G6 repræsenterer, hvordan den danske befolkning så ud pr. 01-01-2023 og placerer husstandene i 100x100 kvadratmeters celler (Geomatic, 2023).

Hver type og gruppe er associeret med variabler, som er aggregeret på decilniveau eller er af typen boolean. Disse variabler dækker eks. over boligtype, civilstand, bilejerskab, familiestørrelse, økonomi, fritidsbeskæftigelser, uddannelsesniveau m.m. Oversigten kan findes på ConZooms hjemmeside.

FynBus høster data i en stor del af busflåden, hvor der indsamles rejsekort data og kameratællinger. Disse data indsamles på rute, tur og stoppestedsniveau. De anvendte påstigningsdata er på stoppested- og månedsniveau for 2023. Datasættet består af data for alle stoppesteder, dvs. letbane- og togstationer er også inkluderet, såfremt der har været registrerede påstigninger i 2023.

Formålet med dette datasæt er at forsøge at anvende det i en regressionsanalyse for at generere nye BKV'er. Disse værdier indgår i beregningsformlen for BKP og bruges til at kvantificere hver enkelt Conzoom-gruppe behov for og/eller villighed til at benytte kollektiv trafik. Med andre ord estimerer BKV sandsynligheden for, at hver Conzoom-gruppe er potentiel kunde i kollektiv transport. For at opnå de mest præcise værdier er alle relevante variabler, der forventes at påvirke brugen af kollektiv trafik, inkluderet i analysen.

Metode

Busværdien er fastsat baseret på gruppen og typernes forbrugsadfærd, som er udarbejdet af Geomatic, Movia og FynBus. BKP defineres ud fra følgende formel:

$$\text{Buskundepotentiale} = \text{Fra Potentiale} + \text{Til Potentiale} \quad (1)$$

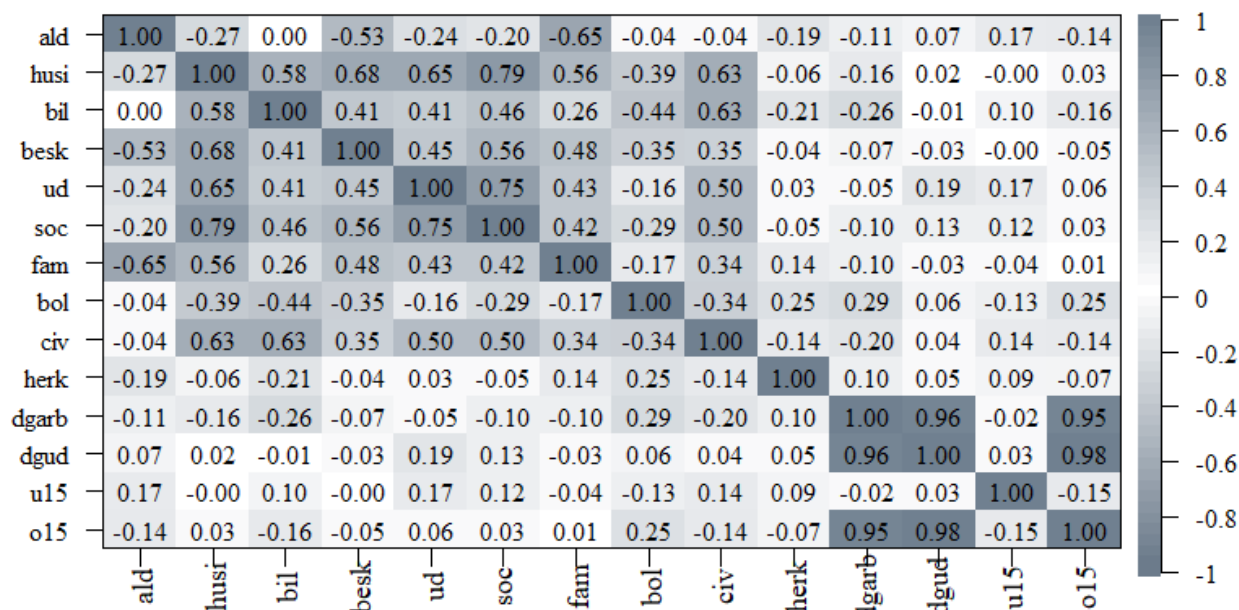
$$\text{Fra Potentiale} = (\text{Natbefolkning}_{\text{AntalPersoner}}) \cdot (\text{BusVærdi FynBus}) / 100 \quad (2)$$

$$\text{Til Potentiale} = (\text{Dagbefolkning}_{\text{Beskaeftigede}} \cdot 0,2) + (\text{Dagbefolkning}_{\text{UddannelseOver15}} \cdot 0,52) \quad (3)$$

FynBus har i dette forsøg valgt at benytte spatiel regression til at forsøge at definere en ny BKV. I første omgang testes metoden på conzoom grupper, fremfor typer grundet begrænsede tidsressourcer.

Passagertallene tager udgangspunkt i historiske data fra 2023, hvor der er benyttet data fra rejsekortet, passagertælleudstyr i busserne og data fra FynBus øvrige salgskanaler

De udvalgte decilvariabler blev analyseret i en indledende undersøgelse, hvor de korrelerede variabler blev udvalgt til videre benyttelse i modellen. Resultaterne blev samlet i en Spearman-korrelationsmatrice for at identificere og ekskludere variabler med høj indbyrdes korrelation. Datasættet afgrænset til kun at bestå af de celler, hvor der har været påstigere/pts-aktivitet, resten er sorteret fra. Resultatet er som følger:



Figur 1: Forkortelser i korrelationsmatricen. ald: aldersdecil, husi: husstandsindkomstdecil, bil: bilr dighedsdecil, besk: besk ftigelsesdecil, ud: uddannelsesdecil, soc: socialklassesdecil, fam: familietyperdecil, bol: boligtypefaktor, civ: civilstandsdecil, herk: herkomstfaktor, dgarb: dagbefolkning i arbejde, dgud: dagbefolkning i uddannelse, uu15: uddannelse under 15, uo15: uddannelse over 15.

Ud fra ovenst ende matrice er decilvariablerne alder, husstandsindkomst, bilr dighed, besk ftigelse, uddannelse, socialklasse, familietype, boligtype, herkomst og civilstand anvendt i regressionsmodellen. Efterf lgende har regressionskoefficienterne for familietyperdecil, herkomst og boligtype vist sig ikke at v re signifikante, hvorfor de er blevet taget ud i den endelige model.

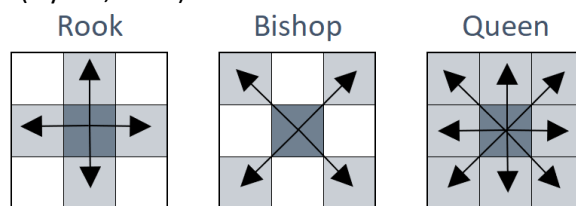
Der blev foretaget fors g med flere spatielle modeller, Spatiel Lag Model, Spatiel Durbin model og Spatiel Error Model. P  baggrund af disse fors g blev Spatiel Error modellen udvalgt grundet laveste AIC-score og en st rre andel signifikante regressionskoefficienter.

Spatiel Error Modellen (SEM) er kendetegnet ved at der tages h jde for korrelationer i fejleddet, der kan v re til stede p  grund af udeladte variabler.

$$y = X\beta + u, \quad \mu = \lambda Wu + \varepsilon \quad (4)$$

hvor y er den afh ngige variabel, X er den uafh ngige variabel, β er koefficienterne, λ er det spatielle fejleddet, som repr senterer graden af den spatielle korrelation i fejleddet, W er den spatielle v gtmatrice med forholdet mellem cellerne og u er en v gtmatrice af uafh ngige fejleddet. (Anselin, 1988).

For at udf re en spatiel regression skal der i modellen defineres afgr nsning i form af nabolag, der spatielle regression modeller som udgangspunkt en af tre typer af nabolag. Disse, baseret p  skakbrikker, er kendt som "Rook", "Bishop" og "Queen". (Oyana, 2021).



Figur 2: Illustration af de tre forskellige nabotyper, der kan v lges mellem

Det vurderes at nabolaget Queen er bedst egnet til regression analysens formål. Dog er der i ConZoom™ observationer som ikke har nogen nabo, fx i kantområder eller isolerede områder. For at undgå at modellen ikke udelukker disse celler er der i modellen specificeret at disse celler skal medtages

Den endelige model defineres derfor som følgende:

$$\begin{aligned} \log(\text{BusPåstigere}) &= \beta_0 + \beta_1 \text{alder} + \beta_2 \text{civilstand} + \beta_3 \text{boligtype} + \beta_4 \text{husstandsindkomst} \\ &+ \beta_5 \text{bilrådighed} + \beta_6 \text{beskæftigelse} + \beta_7 \text{uddannelse} + \beta_8 \text{socialklasse} \\ &+ \text{Conzoom grupper} + \varepsilon \end{aligned} \quad (5)$$

Regressionskoefficienterne som skabes af denne model skal efterfølgende standardiseres til en værdi der kan indgå i beregningen for BKP. Formlen for BKV er designet til at rumme værdier som er på hele tal og på en skala fra 0 – 100. For at kunne indsætte vores nye værdier ind i den eksisterende formel, er det besluttet at standardisere resultaterne så de tilpasses denne skala. Resultaterne fra regressionsanalysen er derfor blevet omregnet til en vægt ved følgende formel:

$$Vægt = \frac{\text{koefficient} - \text{koefficient}_{\min}}{\text{koefficient}_{\max} - \text{koefficient}_{\min}} \Leftrightarrow \frac{\text{koefficient} - -0,85}{2,06 - -0,85} \quad (6)$$

Vægtene er efterfølgende ganget med 100 og afrundet til nærmeste hele tier, så de kan anvendes som BKV i formelen "Fra Potentiale". I nedenstående tabel fremgår de beregnede BKV. Buskundeværdien for Gruppe A. C og F er justeret på baggrund af tidligere model, og vurdering fra Markedsafdelingen.

Resultater

Analysen af den spatielle regression på buskundeværdi har resulteret i koefficienter som er vægtet for at give os en heltalsværdi som der kan benyttes i den eksisterende formel for buskundepotentiale. Denne nye værdi erstatter den forrige, hvor grundlaget for værdien er ukendt, hvilket fremgår af nedenstående Tabel 1.

For at kunne sammenligne forsøget, som er dannet på baggrund af gruppeniveau, med buskundeværdierne fra den eksisterende model, som er dannet på typeniveau, fremgår middelværdien for BKV for gruppen. Dette gøres for at vurderer om værdierne udskiller sig væsentligt fra den tidligere model.

Gruppe kode	Gruppe navn	Koefficient	Vægt	Afrundet BKV	Middel af BKV i dag
A	Velfunderede husejere	-0,38	0,16	20	8
B	Komfort og hygge	0,24	0,38	40	16
C	Lever på landet	-0,85	0,00	5	5
D	Velhavere	0,08	0,32	30	10
E	Urban mangfoldighed	0,38	0,42	40	58
F	Vid og velstand	2,06	1,00	90	36
G	Seniorer	0,44	0,44	40	16
H	Unge på vej	0,38	0,42	40	80
I	Provinsliv	0,53	0,47	50	35

Tabel 1: Resultater for Spatiel Error Regression på ConZoom™ gruppeniveau. Kolonnen "Middel af BKV i dag" er den middelværdien af buskundeværdierne på gruppeniveau.

Da grundlaget er skiftet fra type til gruppe, er der selvfølgelig færre nuancer i buskundeværdien end der var tidligere, hvilket middelværdien også giver udtryk for. Dog ved test af dataene er Buskundepotentialet på det spatielle niveau tilsvarende til den eksisterende model. Derfor er det vurderet at metoden er brugbar i det videre arbejde med buskundepotentiale hos FynBus.

Der vurderes desuden undersøgt om rejsehjemmelfordelinger skal benyttes til at bestemme faktorerne der anvendes i formen Til-Potentiale (3). Der er dog opstillet fire scenarier der kan visualisere hvordan disse faktorer påvirker det endelige buskundepotentiale. I disse scenarier indgår en følsomhedsanalyse hvor vi har inddelt faktorerne. Et hvor de oprindelige faktorer er nedskaleret med henholdsvis 5 og 10 procentpoint. Samt et scenarie som kigger på andelen af beskæftigede, og folk under uddannelse, ud af den samlede befolkning.

Scenarierne tager derfor udgangspunkt i den oprindelige models afgrænsninger hvor det er antaget at 20% af alle i beskæftigelse er potentielle kunder og 52% af uddannelsessøgende over 15 er potentielle kunder

Scenarie 1: tager udgangspunkt i modellens tidligere værdier, hvor 20% af hele dagbefolkningen er potentielle kunder og 52% af alle studiesøgende er potentielle kunder.

Scenarie 2: tager udgangspunkt i scenarie 1, men nedskriver begge værdier med 5 procentpoint.

Scenarie 3: tager udgangspunkt i scenarie 1, men nedskriver begge værdier med 10 procentpoint.

Scenarie 4: tager udgangspunkt i en optimistisk skalering som tager udgangspunkt i det totale potentiale af hele Fyns befolkning, det er antaget denne model at alle beskæftigede og alle under uddannelse er potentielle kunder. Dette indeholder værdierne 43% for dagbefolkningen og 12% af befolkning i uddannelse over 15.

Nedenstående Tabel 2 viser umiddelbare resultater fra følsomhedsanalysen.

Scenarie	Til potentiale	Dagbefolkning i beskæftigelse	Dagbefolkning i uddannelse over 15 år	BKP	Ændring i %
1	20% og 52%	65.494	31.473	231.715	38%
2	15% og 47%	52.619	28.447	215.814	28%
3	10% og 42%	39.744	25.421	199.913	19%
4	43% og 12%	97.235	7.360	239.343	42%

Tabel 2: I ovenstående tabel fremgår resultaterne fra følsomhedsanalysen. BKP står for buskundepotentiale.

Scenarie 1, 2 og 3 vurderes at kunne arbejdes videre med som grundlag for den opdaterede buskundemodell, alt efter interne målsætninger hos FynBus.

Konklusion

Modellen, der anvender spatial regressionsanalyse, viser lovende resultater. Det er dog nødvendigt at undersøge modellens resultater på typeniveau for at sikre, at nuancerne i data er tilstrækkeligt belyst. Derfor fortsættes arbejdet med fokus på at afdække disse usikkerheder, inden modellen implementeres som et analyseværktøj hos FynBus.

Litteratur

Anselin L. 1988. *Spatial Econometrics: Methods and Models*. Kluwer Academics, Dordrecht

Oyana, T. J. 2021. *Spatial analysis with R: Statistics, visualization, and computational methods* (2nd ed). CRC Press.
<https://doi.org/10.1201/9781003021643>

Geomatic. 2023. *De 9 conzoom grupper & 39 conzoom typer*. <https://www.conzoom.dk/da/danmark/segmenter>