



**TURUN
YLIOPISTO**
Kauppakorkeakoulu

Kaukopenelöinnin taustatekijät ja kesto

Rekisteriaineistoon sovellettujen probit- ja survival- mallien tarkastelu

Taloustieteen pro gradu -tutkielma

Laatija:

Samu Kantola

Ohjaaja:

Professori Timo Kuosmanen

4.6.2024

Turku

Turun yliopiston laatujärjestelmän mukaisesti tämän julkaisun alkuperäisyys on tarkastettu Turnitin OriginalityCheck -järjestelmällä.

Pro gradu -tutkielma

Oppiaine: Taloustiede

Tekijä(t): Samu Kantola

Otsikko: Kaukopenelöinnin taustatekijät ja kesto: Rekisteriaineistoon sovellettujen probit- ja survival- mallien tarkastelu

Ohjaaja(t): Professori Timo Kuosmanen

Sivumäärä: 62 sivua

Päivämäärä: 4.6.2024

Perinteisesti liikennetutkimusta on tehty erilaisilla kyselytutkimuksilla. Rekisteriaineiston avulla tehtyjä tutkimuksia penelöinnistä ei juurikaan ole tehty. Tässä tutkielmassa hyödynnetään Tilastokeskuksen rekisteriaineistoa kaukopenelöinnin tutkimiseen. Kaukopenelöinti määritellään pitkäksi matkaksi työpaikan ja kotipaikan välillä. Menetelminä ovat probit, ordered probit sekä survival- mallinnuksessa Coxin suhteellisten vaarojen malli. Tarkoituksena on pyrkiä etsimään taustatekijöitä, jotka selittävät kaukopenelöintiä, kaukopenelöinnin alkamista, sekä kaukopenelöinnin lopettamista. Survival- malleilla pyritään etsimään taustatekijöitä sille, mitkä taustatekijät ovat yhteydessä kaukopenelöinnin lopettamiseen ajassa. Kaukopenelöinnin taustatekijöissä korostuivat erityisesti tutkijakoulutetut, miehet sekä eri kuntaan muuttaneet

Avainsanat: kaukopenelöinti, penelöinti, työmatka, probit

SISÄLLYS

1	Johdanto	7
2	Kaukopenelöinnin käsite	10
	2.1 Liikenneaineistot ja niiden perusteella tehty tutkimus	10
	2.2 Kaukopenelöinnin määrittely	10
3	Probit ja Survival mallit	13
	3.1 Probit- malli	13
	3.2 Survival mallit	14
4	Aineisto ja kuvailevat tunnusluvut	18
	4.1 Aineisto	18
	4.2 Deskriptiivinen tarkastelu	19
	4.3 Deskriptiivinen tarkastelu penelöinnin kestosta	22
5	Kaukopenelöintiä selittäviä tekijöitä	24
	5.1 Probit- mallien tuloksia	24
	5.2 Ordered probit- mallin tuloksia	27
6	Penelöinnin laukaisevia tekijöitä	33
	6.1 Muutokset malliin	33
	6.2 Penelöinnin aloitusta selittävien probit- mallien tuloksia	33
	6.3 Penelöinnin aloitusta selittävän ordered probit- mallin tuloksia	36
	6.4 Penelöinnin lopettamista selittävien probit- mallien tuloksia	38
	6.5 Penelöinnin lopettamista selittävän ordered probit- mallin tuloksia	41
7	Penelöinnin kestoja selittävien mallien tuloksia	44
	7.1 Coxin mallin tuloksia	44
	7.2 Coxin mallin oletusten tarkastelua	46
8	Yhteenveto	56
	Lähteet	59

KUVIOT

Kuvio 1. Helsinkiin suuntautuvan työmatkan kehitys vuosina 2005, 2010 ja 2015	
21	
Kuvio 2. Turkuun suuntautuvan työmatkan kehitys vuosina 2005, 2010 ja 2015	
21	
Kuvio 3. Muutamiin kuntiin kohdistuvien työmatkojen tiheysfunktioiden jakauma matkan mukaan	22
Kuvio 4. Kaplan-Meier käyrät 60 kilometrin selviytymisaineistossa muuttujittain.	
48	
Kuvio 5. Kaplan-Meier käyrät 100 kilometrin selviytymisaineistossa muuttujittain	
49	
Kuvio 6. 60 kilometrin selviytymisaineiston log-log käyrät muuttujittain	53
Kuvio 7. 100 kilometrin selviytymisaineiston log-log käyrät muuttujittain	54

TAULUKOT

Taulukko 1. Käytettävien muuttujien keskiarvo ja keskihajonta	20
Taulukko 2. 60km kaukopendelöijien pendelöinnin yhteenlaskettujen kestojen keskiarvot	23
Taulukko 3. 100km kaukopendelöijien yhteenlaskettujen kestojen keskiarvot	23
Taulukko 4. Yli 100 kilometrin työmatkoja selittävän probit- mallin tulokset	25
Taulukko 5. Yli 60 kilometrin työmatkoja selittävän probit- mallin tulokset	27
Taulukko 6. Aineiston frekvenssit valituilla työmatkojen kynnysväleillä	28
Taulukko 7. Ordered probit- mallin tuloksia	29
Taulukko 8. Tiivistelmä probit- sekä ordered probit- mallien tuloksista	31
Taulukko 9. Probit- mallin tulokset työmatkan periodittaisen muutoksen ollessa yli 100 kilometriä	35
Taulukko 10. Probit- mallin tulokset työmatkan periodittaisen muutoksen kynnyksarvon ollessa 60 kilometriä	36
Taulukko 11. Ordered probit- mallin frekvenssit valituille kynnysväleille pendelöinnin aloittajissa	37
Taulukko 12. Ordered probit- mallin tuloksia pendelöinnin aloittavista	38
Taulukko 13. Probit- mallin tuloksia työmatkan periodittaisen muutoksen ollessa -100km	40
Taulukko 14. Probit- mallin tuloksia työmatkan periodittaisen muutoksen ollessa -60 kilometriä	41
Taulukko 15. Aineiston jakautuminen eri kynnysväleille	42
Taulukko 16. Ordered probit- mallin tuloksia pendelöinnin vähentäjistä	43
Taulukko 17. Cox- mallin vaaran asteen estimaatit 60 kilometrin kynnyksarvolla	
45	

Taulukko 18. Cox mallin vaaran asteen estimaatit 100 kilometrin kynnyksarvolla	
46	
Taulukko 19. Logrank- testin tulokset	51
Taulukko 20. Schoenfeld- testin tulokset 60kilometrin sekä 100 km mallille muuttujittain	55

1 Johdanto

Pendelöinnillä¹ tarkoitetaan edestakaisin työpaikan ja kodin välillä tapahtuvaa säännöllistä liikennettä. Nykyisin tietoverkkojen mahdollistama etätö on hämärtänyt pendelöinnin yhteyttä henkilön fyysiseen siirtymiseen paikasta toiseen. Säännöllinen fyysinen siirtyminen paikasta toiseen ei ole enää välttämättä päivittäistä, vaan voi tapahtua esimerkiksi kerran viikossa. Muina päivinä pendelöinti voi tapahtua tietoverkkoja pitkin. Tässä tutkielmassa pendelöinnin määrittelyyn on sisällytetty myös tämän kaltainen digitaalinen pendelöinti. Rekisteriaineistosta on tämän tutkielman aineiston raameissa mahdotonta selvittää, kuka pendelöijistä on fyysisesti siirtyvä ja kuka tekee etätöitä.

Tarkastelun kannalta tällä kuitenkin nähdään vähäinen merkitys. Työssäkäynnissä merkityksellistä on jonkin alueen työmarkkinalle osallistuminen. Itse fyysinen siirtyminen ei siten ole itsessään lisäarvoa tuovaa, vaan joissain tapauksissa työmarkkinoille osallistumisen mahdollistaja.

Pendelöinnin määrä on ollut yleisesti kasvussa ja työmatkat ovat pidentyneet. Suomessa Tilastokeskuksen Keva (2020) mukaan autoillen kuljetut työmatkat ovat pidentyneet vuosien 2013 ja 2017 välillä noin 900 metriä eli 5%. Ylen (2015) uutisen mukaan vuodesta 1990 vuoteen 2015 saakka viidellä kilometrillä. Yhdysvalloissa vastaavasta ilmiöstä on uutisoitunut U.S Census Bureau (Burd ym., 2021). Pendelöintiin liittyvät teemat ovat olleet myös lähivuosina pinnalla taloustieteen huippujulkaisuissa. Esimerkiksi Monte ym. (2018) löysi pendelöintikustannusten vähentämisestä hyvinvointivaikutuksia. Owens ym. (2020) löysi yleisemmin kaupunkisuunnittelusta suuria tehottomuuksia pendelöintikustannusten kannalta tarkasteltuna.

Pendelöinnillä on eräitä tärkeitä tarkoituksia yhteiskunnassa. Se tasaa alueellisia työllisyyseroja, koska korkean työttömyyden alueilta on mahdollista pendelöidä matalan työttömyyden alueelle töihin. Samoin pendelöinti tasoittaa paikkakohtaisia pullonkauloja esimerkiksi asumisessa. Pendelöinnin on myös havaittu tasoittavat

¹ Pendelöinti- sana tulee latinan sanasta pendulum, mikä tarkoittaa heiluria. Pendelöinnissä ihmiset työaikojen mukaisesti tasaisin väliajoin ”heilauttavat” heilurin tapaan työpaikan ja kodin välillä tehden samaa liikettä päivästä toiseen. Pendelöinti, eli työmatkojen kulkeminen, on siten lähes jokaista koskettava ilmiö yhteiskunnassa, ja siten tutkimuksen arvoinen.

alueellisia tuloeroja ja toimivan tulojensiirtomekanismina korkean työpaikkatiheyden alueilta maaseudulle. Pendelöinnillä on siten myös rakennemuutostilanteissa merkittävä rooli rakennemuutoksia tasoittavana toimintana. (Holm ym. 2008, 55.)

Erityisen kiinnostavana ryhmänä pendelöijistä ovat kaukopendelöijät, jotka käyvät töissä erittäin pitkänkin välimatkan takaa. Kaukopendelöinnin tutkimuksen tekee kiinnostavaksi sen trendinomaisuus, sekä sen mahdolliset ongelmat yhteiskunnassa. Kahneman & Krueger (2006) kyselytutkimuksessa työmatkat olivat itse työskentelyn kanssa eniten negatiivisia tunnetiloja aiheuttanut asia. Samantyyppisiä tuloksia on Ford Motor Companyn (2015) teettämässä kyselytutkimuksessa, missä työmatka oli jopa stressaavampi osa päivästä kuin itse työpäivä. Sosiaalista pääomaa Harvardissa tutkinut Robert Putnam on todennut Paumgartenin kirjoittamassa New Yorkerin artikkelissa (2007) nyrkkisääntönä jokaisen kymmenen minuutin lisäaika työmatkaan vähentää sosiaalisia suhteita kymmenellä prosentilla. Kaukopendelöinnin voidaan siis sanoa olevan keskustelua herättävä aihe yhteiskunnassa.

Tämä tutkielma pyrkii perehtymään kaukopendelöijiin ryhmänä. Ekonometrisiä malleja käyttäen pyritään identifioimaan millaiset taustatekijät nostavat todennäköisyyttä sille, että henkilö on kaukopendelöijä. Toisena tavoitteena on selvittää, millaiset taustatekijät voisivat toimia kaukopendelöinnin aloittamisen ”liipasimena”. Tutkimuksessa ei pyritä selvittämään minkäänlaista kausaalisuutta taustatekijöiden ja pendelöinnin välillä.

Monet aikaisemmat pendelöintiä sekä kaukopendelöijiä käsittelevät tutkimukset ovat tehty kyselytutkimuksilla ihmisten pendelöintitottumuksista. Tässä tutkielmassa kyselytutkimuksen sijaan käytetään Tilastokeskuksen tarjoamaa rekisteriaineistoa. Rekisteriaineiston eduksi voidaan laskea suuri otoskoko, sekä sen epävarmuuden poistuminen, vastaavatko ihmiset todenmukaisesti kyselyihin.

Suomessa aikaisempaa tutkimusta edustaa Holm ym. (2008). Tutkimus on sikäli erilainen tyypilliseen aiheen tutkimukseen verrattuna, että siinä on käytetty rekisteriaineistoa. Holm ym. (2008) tutki muun muassa kuka pendelöi ja kuka muuttaa toiselle paikkakunnalle.

Tämä tutkielma pyrkii vastaamaan seuraaviin kysymyksiin:

1. Mitkä sosioekonomiset ja muut taustatekijät selittävät kaukopendelöintiä?

2. Mitkä taustatekijät voivat laukaista tai lopettaa kaukopenelöinnin?
3. Mitkä taustatekijät toimivat kaukopenelöintiä ylläpitävänä tekijänä?

Kaukopenelöinnin taustatekijöitä, sekä kaukopenelöinnin laukaisevia ja lopettavia taustatekijöitä pyritään selvittämään probit sekä ordered probit- malleilla. Pendelöinnin ylläpitäviä tekijöitä kartoitetaan survival- mallilla. Näiden mallien avulla on mahdollista luoda kaukopenelöijän profiili, joka tuo arvokasta tietoa siitä, millaisille ihmisille kaukopenelöinnin hyödyt ja haitat kohdistuvat ja millaiset ihmiset hyötyvät mahdollisista uusista liikennehankkeista, esimerkiksi nyt pinnalla olleesta Turun ja Helsingin välisestä Tunnin juna- hankkeesta.

2 Kaukopenselöinnin käsite

2.1 Liikenneaineistot ja niiden perusteella tehty tutkimus

Työmatkoista on koottuna kansainvälisesti useita erilaisia aineistoja, missä kerätään tietoa kansalaisten työmatka- tottumuksista. Näitä aineistoja yhdistävä tekijä on se, että ne ovat kerätty kyselytutkimuksella. Yhdysvalloissa on muun muassa American Community Survey, ja Euroopassa European long-distance travel mobility survey. Näiden lisäksi on myös pienempiä kansallisesti kerättyjä aineistoja, kuten Suomessa Trafín tekemä Valtakunnallinen henkilöliikennetutkimus tai Ruotsin valtion alaisen Trafik analysin tekemä Resvanor i Sverige- tutkimus.

Kyselytutkimukset ovat siis eräänlainen alan standardi selvittämään työmatka- tottumuksia. Tässä tutkielmassa uutta on Tilastokeskuksen rekisteriaineiston käyttö.

2.2 Kaukopenselöinnin määrittely

Kaukopenselöinnin tai pitkän työmatkan (engl. *extreme commuting*) määrittely on ollut ongelmallista maiden erilaisten ominaisuuksien vuoksi. Kaupungistumisaste, sekä kaupunkien koko on eräs suuri maita erottava tekijä. Myös menetelmät ovat olleet moninaisia. Selvyyttä ei ole edes siitä, pitäisikö työmatkaa tarkastella aikana vai matkan pituutena.

Eräs yleinen tapa on mielivaltaisesti asettaa kynnsarvo johonkin tiettyyn kohtaan. Yhdysvaltojen kontekstissa ajan kautta aihetta lähestyvät Marion & Horner (2007) asettavat rajan yksisuuntaiseen 90 minuutin matkaan. Dana (2011) määrittää pitkäksi työmatkaksi 90 minuuttia tai enemmän päivittäin kestävän matkustamisen keskiverto työssäkävyn työmatkan ollessa alle 30 minuuttia.

Matkakohtaisia mielivaltaisia kynnsarvoja on lukuisia eri maakohtaisia Yhdysvaltojen sadasta mailista (163 kilometriä) Yhdistyneen kuningaskunnan 50mailiin (83 kilometriä). Suomen osalta kiinnostava arvo on Ruotsin Trafik Analysin määrittämä 100 kilometrin kynnsarvo. (Limtanakool ym. 2006, 390.)

Limtanakool ym. (2006) pyrkivät lähestymään tätä toisella tavalla kuin vain asettamalla mielivaltaisen kynnsarvon. He asettivat kynnsarvon sen mukaan, mikä on

keskimääräinen metropolialueiden koko. Tällä menetelmällä Yhdistyneen kuningaskunnan sekä Alankomaiden osalta 50 kilometriä tuli pitkän työmatkan kynnsarvoksi.

Tämä tarkastelu ei sovellu määrittämään kovinkaan hyvin Helsingin ja Turun väliseen tilanteeseen, sillä Helsingin keskustasta etäisyys pääkaupunkiseutuna pidetyn Espoon tai Vantaan ulkorajalle on noin 20 kilometriä. Turun tapauksessa vastaava etäisyys Kaarinan tai Raision ulkorajoille olisi noin 10 kilometriä, ja siinäkin olisi kiistanalaista, onko kyse yhtenäisestä alueesta. Toisaalta, myös hyvin matalia kynnsarvoja on käytetty määrittämään pitkää työmatkaa. Maoh & Tang (2012) esimerkiksi asettivat kynnsarvoksi 17 kilometriä tutkiessaan Kanadan Windsoriin suuntautuvia työmatkoja ja Champion ym. (2009) asettivat Englannin maaseudulla pitkän työmatkan arvoksi 20 kilometriä.

Tämän tutkielman osalta Ruotsin Trafik Analysin sadan kilometrin kynnsarvo on mielivaltaisista arvioista käyttökelpoisin Suomea koskien pitkän työmatkan määrittelemiseen. Tämän arvon käyttökelpoisuutta voi perustella Helminen & Ristimäki (2007) tuloksilla. Mikäli henkilöllä on yli sata kilometriä matkaa kodin ja työpaikan välillä, työpaikan lähellä oleva asunto nousi tutkimuksen tekemisen ajankohtana yleisimmäksi tavaksi välttää työmatka. Siten sata kilometriä sopii tuolta osin hyväksi kynnsarvoksi pitkälle työmatkalle.

Aikaisemmassa tutkimuksessa korostuvat lähes kaikissa demografiset sekä sosioekonomiset selittäjät. Marion & Horner (2007) on kattavasti sosioekonomisiin sekä demografisiin tekijöihin perehtyvä tutkimus, minkä motivaationa on ollut sosiaalisen oikeudenmukaisuuden painotus ja epätasa-arvo- näkökohdat. He ovat käyttäneet pitkän työmatkan selittäjänä ikää, sukupuolta, koulutustasoa, siviilisäätyä, lapsia, rotua, tuloja, yhteiskäyttöautoa (engl. *car pooling*), asunnon omistajuus, työmatkalle lähtöaika, työaika sekä alueelle muutto ajankohta. Limtanakool ym. (2006) edustaa yleisempää linjaa selittäjien valinnassa. He eivät käytä yhtä kattavasti sosioekonomisia ja demografisia selittäjiä, mutta käyttävät väestötiheyttä yhtenä selittäjänä sekä mahdollisuutta käyttää autoa. Vastaavaa linjaa pääpiirteissään edustaa myös Geslin & Ravaet (2016), Ricard (1988) ja Maoh & Tang (2012) ja Champion (2009).

Selitysarvoa pitkän työmatkan kulkemiseen on löytynyt koulutustasosta, kimppakyydin mahdollisuudesta, sukupuolesta sekä tulotasosta (Marion & Horner 2007, 41),

väestötiheyden tasaisuus (Limtanakool ym. 2006, 402), muuttaja- status (Champion ym. 2009, 1256-1257), asuminen metropoli- alueella, hyvät junayhteydet, ankkuroituminen tietylle alueelle (Geslin & Ravalet 2016, 243-245), mahdollisuus etätööhön (Helminen & Ristimäki 2007, 339), ammatti, ikä, sekä kokoaika työ (Maoh & Tang 2012, 54).

Liikenneministeriö on tehnyt Suomesta selvityksen, jossa vertailtiin työssäkävijän valintaa pendelöinnin ja työpaikkakunnalle muuttamisen välillä. Asuinseudun ulkopuolella olevat korkeammat palkat, mies- sukupuoli, vanha ikä ja matala koulutus ennustivat liikkuvuutta (Alasalmi ym. 2020, 83). Tämä voisi olla yhtenevä Geslin & Ravaletin (2016) löytämään ankkuroitumisvaikutukseen tietylle seudulle. Muuttamista ennustivat nais- sukupuoli, nuoruus, korkea koulutus, hyvä vuokra-asuntojen saatavuus sekä matala vuokrataso. Asuntomarkkinat olivat ylipäättään suurempana erottajana mitä työmarkkinat siinä, valitseeko työssäkävijä pendelöinnin vai paikkakunnalle muuttamisen. Erityisen alhaisina liikkujina ovat pääkaupunkiseudulla asuvat, mihin vaikuttavat todennäköisesti pääkaupunkiseudun laajat työmarkkinat (Alasalmi ym. 2020, 83-84). Holm ym. (2008) mukaan muuttajia todennäköisesti ovat työvoiman ulkopuolella olevat ja työttömät sekä nuoret. Miesten ja naisten muuttoalttiudessa ei ole eroja, mutta miehet pendelöivät huomattavasti enemmän.

3 Probit ja Survival mallit

3.1 Probit- malli

Tutkielmassa käytetään kahta probit- sukuista mallia: Probit- mallia sekä Ordered Probit- mallia. Nämä mallit tuntuvat luontevalta, sillä pendelöinnin aloitus tai lopetus on binäärinen- tapahtuma. Probit mallin muoto on seuraava:

$$\pi_{xi} = \Phi \left(\alpha + \sum_{j=1}^n \beta_j x_{ij} \right)$$

missä π_x on selitettävä muuttuja. Probit- mallin erikoisuutena selitettävä muuttuja on ehdollinen todennäköisyys selittävistä muuttujista x . Siten π_x arvon tulee olla suljetulta väliltä $[0,1]$. Φ on normaalijakauman kertymäfunktio ja α sekä β ovat aineiston parametrit, vakiotermin ja regressiokerroin. Voidaan havaita, että mikäli $\beta > 0$, todennäköisyys selitettävälle asialle kasvaa monotonisesti kun x kasvaa, ja päinvastoin. (McCullagh 1989, 31.) Probit- mallissa termi $\alpha + \sum_{j=1}^n \beta_j x_{ij}$ voidaan tulkita Z-arvoksi, joka kertoo, kuinka monen keskihajonnan verran havaittu arvo on poikkeava suhteutettuna keskihajontaan (Stock & Watson 2007, 391).

Ordered probit- malliin otetaan pohjaksi usean muuttujan probit- malli. Selitettävä muuttuja π_{xi} voi nyt saada muitakin arvoja kuin väliltä $(0,1)$. Muuttujan π_{xi} arvot voivat nyt olla seuraavat:

$$\pi_{xi} = \begin{cases} 1 \text{ jos } D_i \leq \delta_1 \\ 2 \text{ jos } \delta_1 \leq D_i \leq \delta_2 \\ \vdots \\ n \text{ jos } \delta_{n-1} \geq D_i \end{cases}$$

missä D_i on tässä tapauksessa erilaisia työmatkoja. aineistosta havaittu järjestysasteikollinen suure, esimerkiksi hyönteismyrkkyesimerkissä hyönteismyrkyn aiheuttama haitan aste hyönteiselle. δ_n on valitut kynnyсарvot, mitkä määrittävät mihin kategoriaan havaittu D_i asettuu. Kynnyсарvojen perusteella siis voidaan analysoida eri pituisia työmatkoja kulkevia ihmisiä eri ryhmissä (Van Kant 2002, 8-9.)

Probit- mallin estimointiin koneellisesti käytetään suurimman uskottavuuden menetelmää. Suurimman uskottavuuden menetelmässä maksimoidaan uskottavuusfunktioita, mistä saadaan suurimman uskottavuuden estimaattori.

Uskottavuusfunktio on aineiston yhteistodennäköisyysjakauma, missä on tuntemattomia parametrejä. Uskottavuusfunktio on yleisessä muodossaan seuraavanlainen:

$$\mathcal{L} = \mathcal{L}(\theta, y, x) = p_y(y, \theta, x).$$

Nyt x ja y ovat havaittuja aineistoja. $\theta \in \Theta$, missä θ on vektori parametriavaruudessa Θ . θ on joukko tuntemattomia vakioparametrejä. (Cramer 1986, 13-14.)

3.2 Survival mallit

Survival- malleja käytetään estimoimaan jonkun havainnoitavan asian selviytymistä ajassa havainnointijaksolla. Yleisesti survival- malleja on käytetty lääketieteessä havainnoimaan esimerkiksi johonkin tautiin jotain hoitoa saaneiden ihmisten selviytymistä ajassa verrokkiryhmään verraten, jotka eivät ole saaneet hoitoa.

Pendelöinnin osalta survival- mallilla voi pyrkiä ennustamaan pendelöinnin kestoa. Tässä tapauksessa aineistosta on havaittavissa jokaisen havainnon pendelöinnin aloitus, lopetus sekä pendelöintiperiodit. Tässä tutkielmassa ensin keskitytään pendelöinnin kestoon ensin yleisesti käyden läpi lyhyesti Nelson-Aalen-, sekä Kaplan-Meier-estimaattorit. Sen jälkeen rakennetaan Coxin suhteellisten vaarojen malli (myöh. *Coxin malli*).

Nelson-Aalen-, sekä Kaplan-Meier- estimaattorit pyrkivät luomaan estimaatin vaaran asteesta (engl. *hazard rate*). Mielenkiinnon kohteena ovat ajanjaksot ennen tapahtumia (engl. *event*). Näistä ajanjaksoista ennen tapahtumaa on mahdollista päätellä todennäköisyys seuraavalla periodilla tapahtuvalle tapahtumalle, eli estimoida vaaran aste $\alpha(t)$. Koska $\alpha(t)$ muodolle tai ominaisuuksille ei ole juuri mitään muita rajoituksia kuin $\alpha(t) > 0$, on käytännöllisempää estimoida kumulatiivinen vaara

$$A(t) = \int_0^t \alpha(s) ds$$

mistä saadaan Nelson-Aalen- estimaattori

$$\hat{A}(t) = \sum_{T_j \leq t} \frac{1}{Y(T_j)}$$

missä Y on tapahtuman ajanhetkestä T riippuva aineisto. (Aalen ym. 2008, 71-72).

Toisenlainen lähestymistapa on Kaplan-Meier-estimaattori, mikä on enemmän käytetty näistä kahdesta. Kaplan-Meier-estimaattori toimii paremmin kuin Nelson-Aalen-estimaattori, mikäli aineisto on sensuroitua (engl. *censored*). Aineiston sensuroitumisella tarkoitetaan havainnon ”katoamista” havaintoaineistosta ennen havainnoidun ajanjakson päättymistä. Kaplan-Meier-estimaattori perustuu ehdollisten todennäköisyyksien kertolaskusääntöön ja on muodoltaan seuraavanlainen:

$$\hat{S}(t) = \prod_{T_j \leq t} \left\{ 1 - \frac{1}{Y(T_j)} \right\}$$

missä muuttujat ovat samat kuin Nelson-Aalen-estimaattorissa. Mikäli otoksessa ei ole oikealta sensuroituja havaintoja, Kaplan-Meier-estimaattori redusoituu yksi miinus otoksen kumulatiivinen jakaumafunktioksi. (Aalen ym. 2008, 90-91). Kaplan-Meier-estimaattorin, sekä Nelson-Aalen-estimaattorin yhteys on sama kuin selviytymisfunktion (engl. *survival function*) ja kumulatiivisen vaaran asteen (Aalen ym. 2008, 99).

Coxin malli kuuluu suhteellisten vaarojen malleihin (engl. *relative risk regression*), mitkä ovat yleisimpiä survival- malleja. Coxin malli on muodoltaan seuraavanlainen:

$$\alpha_i(t) = \alpha_0(t) e^{\beta^T x_i(t)}$$

missä $x_i(t)$ on selittävien muuttujien vektori, β on regressiokertoimien vektori ja α_i on selitettävä muuttuja. α_0 on mallin ”maailman” vakio vaarallisuuden aste (engl. *baseline hazard*). (Aalen ym. 2008, 133.)

Mallin regressiokertoimet tulkitaan vaaran asteina. Mikäli vaaran aste on $\beta > 1$, vaara on lisääntynyt, eli selviytymistodennäköisyys pienentynyt. Kun vaaran aste $\beta < 1$, tilanne on päinvastainen. Jos regressiokerroin $\beta = 1$, selittäjä ei vaikuta ja selitettävä muuttuja kohtaa vain vakioisen vaaran. (Aalen ym. 2008, 137.)

Niin Kaplan-Meier, kuin Coxin mallikin, pyrkii esittämään samaa asiaa. Coxin mallilla on kuitenkin joitain tämän tutkielman kontekstissa olevia etuja, mitkä Kaplan-Meieriltä puuttuu. Kaplan-Meier-estimaattoria on mahdollista käyttää vain yksinkertaiseen vertailuun tutkimus- sekä verrokkiryhmän välillä, ja näiden vaaran asteen vertailuun.

Coxin malliin on mahdollista lisätä useita eri selittäviä muuttujia ja on siten tässä tutkielmassa mieluisampi.

Coxin mallin puute verrattuna Kaplan-Meier-estimaattoriin on kuitenkin nimensä mukaisesti vahva oletus suhteellisista vaaran asteista. Tämä tarkoittaa, että eri selittäjien vaarojen asteiden on oltava keskenään vakioisia ajan yli eri muuttujien saamalla arvoilla.

Coxin mallin tutkimiseen suhteellisten vaarojen asteista on tarjottu useita eri vaihtoehtoja, missä tässä käytetään kahta visuaalista menetelmää, sekä yhtä tilastollista testiä. Visuaalisista menetelmistä vertailevaa tutkimusta on tehnyt muun muassa Hess (1995). Tässä tutkielmassa käytetään yleisesti käytetyimpiä menetelmiä, mistä visuaaliseen tarkasteluun perustuvat ovat Kaplan-Meier-käyrien tarkastelu, sekä log-log-kuvaajien tarkastelu. Tilastollista testiä edustaa Schoenfeldin jäännöksiin perustuva testaaminen sekä logrank-testi.

Oletusta suhteellisista vaaran asteista on mahdollista tutkia intuitiivisimmillaan tarkastelemalla visuaalisesti Kaplan-Meier-estimaattorin luomia kuvaajia verrokkiryhmästä, sekä tutkimusryhmästä. Suhteellisen vaaran asteen pitäisi siis oletuksen mukaisesti pysyä samana käytettävien muuttujien välillä. Kaplan-Meier-estimaattorissa muuttujien kuvaajien osoittamien arvojen erotuksen tulisi siis pysyä vakiona. Lisäksi yksittäisen muuttujan vaaran asteen tulisi eri ajanhetkillä pysyä vakiona suhteessa edeltäneeseen ajanhetkeen. Tämän menetelmän rajoitteena on Kaplan-Meierin yleinen rajoite tarkastella vain kategorisia muuttujia, mutta ei jatkuvia muuttujia.

Kuvaajassa tämä tarkoittaa sitä, että tutkimusryhmän, sekä verrokkiryhmän Kaplan-Meier-kuvaajien tulisi olla samansuuntaiset koko matkalta. Kuvaajien risteämäkohta voidaan tulkita heti suhteellisten riskien oletusten rikkoontumiseksi. Kuvaajien raju erkaneminen toisistaan on mahdollinen oletuksen rikkoontuminen. Suhteellisten riskien oletusten rikkoontumisen ajankohdalla on myös merkitystä. Mikäli rikkoontuminen tapahtuu käyrän oikeassa hännässä, suurin osa aineistosta on jo käytetty ja rikkoontuminen ei ole yhtä vakava asia.

Kaplan-Meierin kanssa voidaan käyttää muuttujakohtaista logrank-testiä. Logrank-testi testaa, ovatko kaksi Kaplan-Meier kuvaajaa keskenään samanlaiset. Nollahypoteesi logrank-testissä on, että kuvaajien jakaumat ovat samanlaiset. $P < 0,05$ -arvolla

nollahypoteesi hylätään ja jakaumat ovat erilaiset. Tällöin on mahdollista päätellä, että muuttuja on vaikuttanut vaaran asteeseen.

Toinen visuaalinen menetelmä tutkia suhteellisten riskien oletuksen toteutumista on tarkastella log-log- kuvaajia. Tässä menetelmässä Coxin malliin tehdään kaksinkertainen logaritmuunnos

$$\ln[-\ln(\alpha_i(t))]$$

minkä jälkeen tarkastellaan, ovatko näin saadut kuvaajat yhdenmukaisia toistensa kanssa. Log-log- muunnettu Coxin malli on linearisoitu, mikä tarjoaa visuaalisesti helpomman tavan tarkastella kuvaajien eroavaisuuksia muuttujien, sekä ajan välillä.

Log-log- kuvaajan tulkintaan pätee samat säännöt kuin Kaplan-Meier- kuvaajienkin: mikäli kuvaajat muuttujien välillä ovat yhteneväisen näköiset, sekä kuvaajat tutkimus- sekä verrokkiryhmän välillä ovat yhdensuuntaiset. Tämän menetelmän etuna verrattuna Kaplan-Meier- kuvaajan tarkasteluun on soveltuvuus myös jatkuvien muuttujien analysointiin.

Kolmas menetelmä on David Schoenfeldin (1982) kehittämä menetelmä, missä voidaan käyttää Coxin mallin ajasta riippumatonta Schoenfeld- jäännöstä. Schoenfeld- jäännös on erotus tapahtuman kokevan havaintopisteen selittävien muuttujien arvoista ja muiden alttiina tapahtumalle olevien havaintopisteiden riskipainotettu keskiarvo. Schoenfeld- jäännökset ovat siinä mielessä käteviä, että niiden ajassa muuttuvaa trendiä on mahdollista testata χ^2 - testillä. Nollahypoteesi on, että Schoenfeld- jäännökset eivät ole korreloituneita ajan, tai muiden selittävien muuttujien kanssa. Nollahypoteesin hyväksyminen kertoo siitä, että suhteellisen riskin oletus pitää ajassa, eikä siinä ole mitään jaksottaista trendiä, tai muuta muutosta ajassa.

4 Aineisto ja kuvailevat tunnusluvut

4.1 Aineisto

Työmatkan aineistona käytetään Tilastokeskuksen Suomesta keräämää rekisteriaineistoa. Rekisteriaineiston käyttö tuo hyötyjä kyselytutkimukseen nähden. Rekisteriaineistossa on edustettuna koko väestö, jolloin otoskoko on suuri. Otos ei ole myöskään vinoutunut sen perusteella, mikä osa väestöstä vastaa eniten kyselytutkimuksiin. Väestö voidaan tunnistaa aineistosta anonymisoitujen sosiaaliturvatunnusten perusteella ja tutkia pendelöinnin muutoksia ajassa muiden muuttajien muutosten kanssa.

Rekisteriaineiston käyttö asettaa kuitenkin omia rajoitteita verrattuna kyselyaineistoon. Työmatka-aineisto kuvaa työmatkan lineaarista etäisyyttä, eli niin kutsuttua linnuntie-etäisyyttä, työpaikan ja asuinpaikan koordinaattien välillä. Kyselytutkimuksessa saadut omat arviot työmatkasta saattavat paremmin kuvastaa erityisesti lyhyiden työmatkojen tapauksessa todellista työmatkaa. Pidemmällä työmatkalla nämä erot tasoittuvat Suomen valtatieverkoston vuoksi. Siten suurta harhaa ei pidempien työmatkojen, saati kaukopendelöinnin, tapauksessa pitäisi syntyä.

Rekisteriaineisto ei kykene erottelmaan muun kuin työpaikan toimipaikan perusteella henkilöiden työmatkoja. Mahdollinen keikkatyö ja muu, missä työmatka harvoin välttämättä kohdistuu kotoa yrityksen toimipaikkaan, jäävät rekisteriaineistossa huomioimatta.

Käytetty aineisto koostuu vuodesta 2005 vuoteen 2017. Aineistosta on jätetty pois yli 300km työmatkat. Nämä ovat Tilastokeskuksen aineistokatalogin mukaan epäluotettavia. Tutkimus keskittyy Uudenmaan ja Varsinais-Suomen alueelle. Uusimaa ja Varsinais-Suomi työssäkäyntialueena on kiinnostava moneltakin kantilta katsottuna. Alueella on Tunnin juna- hanke Turun ja Helsingin välille, jolla kaukopendelöintiä pyritään helpottamaan. Tämän lisäksi, alueiden välillä on VT1- moottoritie, mikä on ollut pitkäkestoinen hanke parantamaan mahdollisuuksia pendelöidä alueella. Uudenmaan ja Varsinais-Suomen alueella on lisäksi kaksi taloudellisesti merkittävää kasvukeskusta: PK- seutu, sekä Turun seutu.

4.2 Deskriptiivinen tarkastelu

Probit- mallin tarkoituksena on saada tietoa vain kaukopendelöijistä, sekä kaukopendelöinnin laukaisevista tekijöistä. Ordered probit mallilla ryhmäjako on tarkempi ja sen avulla kyetään selittämään kaikenlaista pendelöintiä ja niitä laukaisevia tekijöitä. Probit ja Ordered probit- mallilla ei tässä tutkielmassa kyetä tekemään kausaalipäätelyä, vaan tarkoitus on vain kuvata minkälaiset tekijät toimivat pendelöinnin selittäjinä.

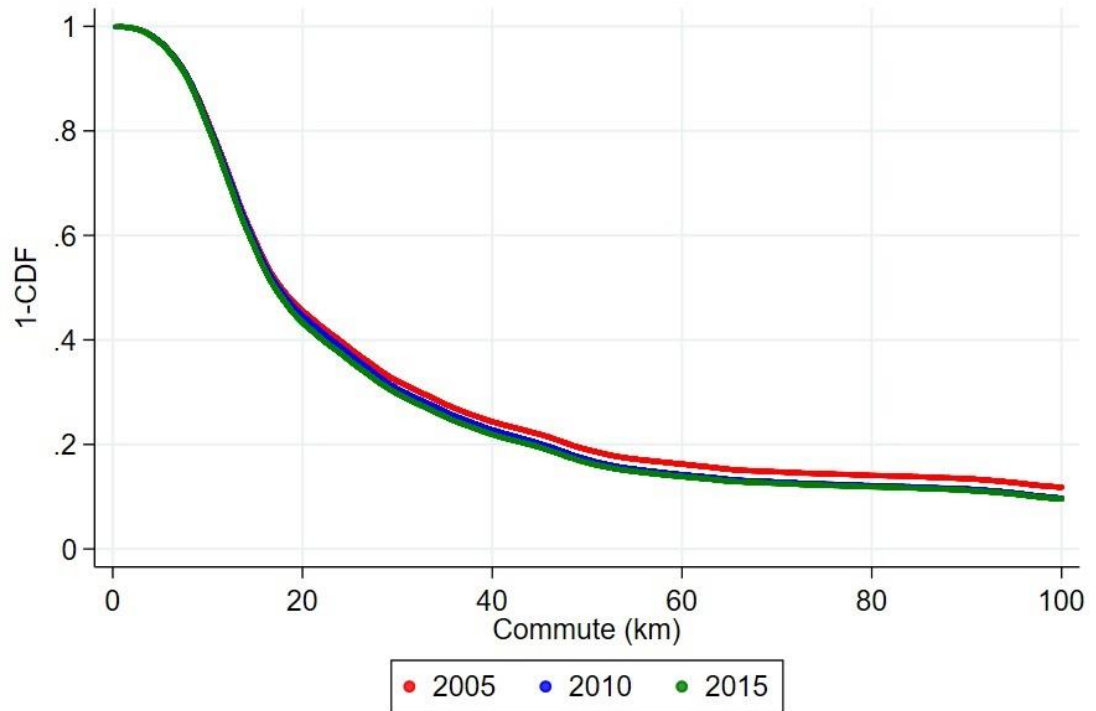
Useat erilaiset tekijät voivat olla oleellisia pendelöinnin selittämisessä. Mallissa pääpaino on henkilökohtaisissa tekijöissä sekä alueellisissa tekijöissä. Henkilökohtaisia tekijöitä ovat muun muassa ikä, pienten lasten lukumäärä ja koulutus. Alueellisia tekijöitä junaradan henkilöpaikka, valtatie ja metropolialue.

Taulukko 1 kertoo käytettävien muuttujien keskiarvon sekä keskihajonnan. Otoksoon 7 027 664 aineistossa keski-ikä on noin 43 vuotta ja 48% on miehiä. Aviossa on yli puolet otoksesta ja vähän yli puolet asuu VT1 tai/ja pääkaupunkiseudulla, mihin kuuluvat Helsinki, Vantaa, Espoo sekä Kauniainen. Työttömyyskuukausia viimeisen vuoden aikana on ollut keskimäärin vajaa 0,06 kuukautta. Muuttajia, joiden kotikunta on muuttunut joidenkin tilastointivuosien välillä, on noin 4% otoksesta ja ulkomaalaistaustaisia on 6,2%. Korkein koulutusaste on toinen aste vajaalla 40% otoksesta, suunnilleen vastaavalla määrällä on jonkunlainen korkeakoulutus tutkijakoulutuksen saaneiden osuuden jäädessä pariin prosenttiin otoksesta.

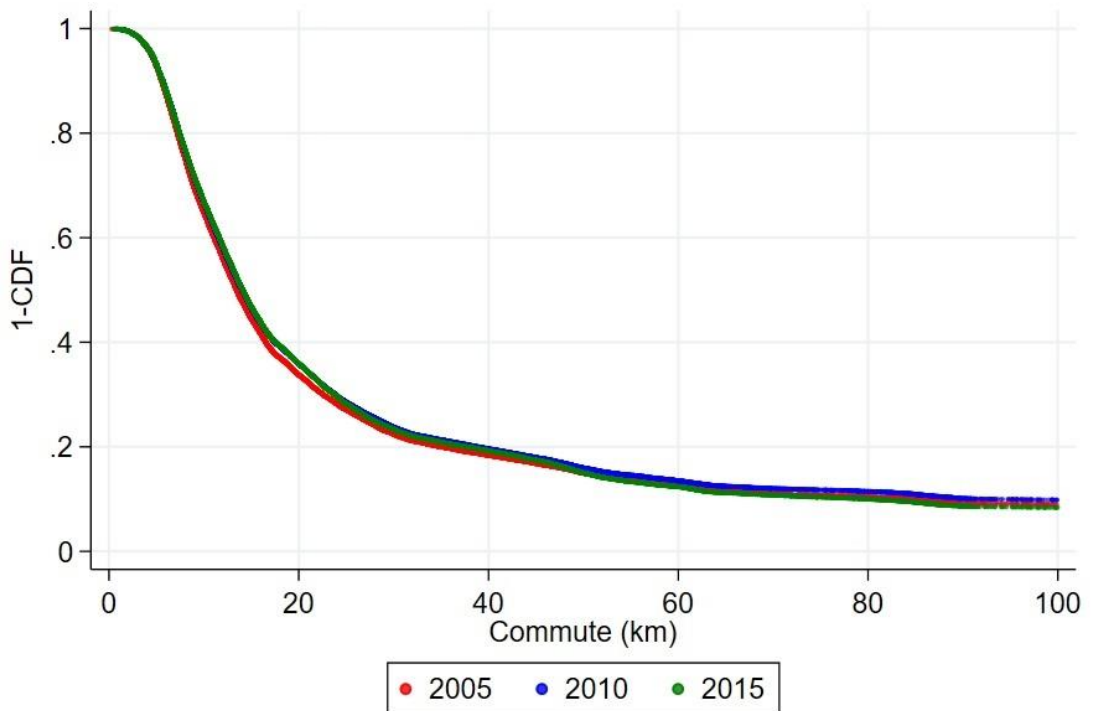
Taulukko 1. Käytettävien muuttujien keskiarvo ja keskihajonta

	Keskiarvo	Keskihajonta
Juokseva vuosi	7.498	3.440
Henkilökohtaiset muuttujat		
Ikä	43.024	11.230
Mies	0.481	0.500
Työttömyyskuukausia vuoden aikana	0.062	0.499
Aviossa	0.641	0.480
Omistusasuja	0.760	0.427
Vuokra-asuja	0.195	0.396
Käytössä olevat tulot tuhansina	31.645	16.631
Alle 7- vuotiaiden lasten lukumäärä	0.334	0.657
Tulojen muutos t-1 ja t välillä	0.000	1.830
Eri kuntaan muuttaja	0.040	0.196
Ylempi toimihenkilö	0.301	0.459
Alempi toimihenkilö	0.384	0.486
Työntekijä	0.223	0.416
Ulkomaalaistausta	0.052	0.223
Korkein tutkinto toinen aste	0.371	0.483
Korkein tutkinto erikoisammattikoulu	0.010	0.098
Korkein tutkinto alin korkea-aste	0.142	0.349
Korkein tutkinto alempi korkeakoulu	0.148	0.355
Korkein tutkinto ylempi korkeakoulu	0.178	0.382
Korkein tutkinto tutkijakoulutus	0.020	0.139
Asuinpaikkakohtaiset muuttujat		
Asuminen pääkaupunkiseudulla	0.516	0.500
Rantarata asuinkunnassa	0.540	0.498
Päärata asuinkunnassa	0.150	0.357
VT1 asuinkunnassa	0.568	0.495

Kuvio 1 esittää pendelöinnin kehityksen Helsinkiin vuosina 2005, 2010 ja 2015. Kuvio 2 esittää pendelöinnin Turkuun samoina vuosina. Kuvioissa 1 ja 2 Y- akselilla on 1-pendelöijien kumulatiivinen tiheysfunktio, sekä X- akselilla kilometrimäärä. Kuvioiden 1 ja 2 tulkinta on, että Y- akselilta on luettavissa se osuus pendelöijistä, mikä pendelöi pidemmälle, kuin vastaava X- akselin osoittama kilometrimäärä on. Pendelöinti on lisääntynyt Helsinkiin välillä 2005-2015, erityisesti pidemmissä työmatkoissa. Turkuun pendelöinnissä vastaavaa muutosta ei ole juurikaan tapahtunut.



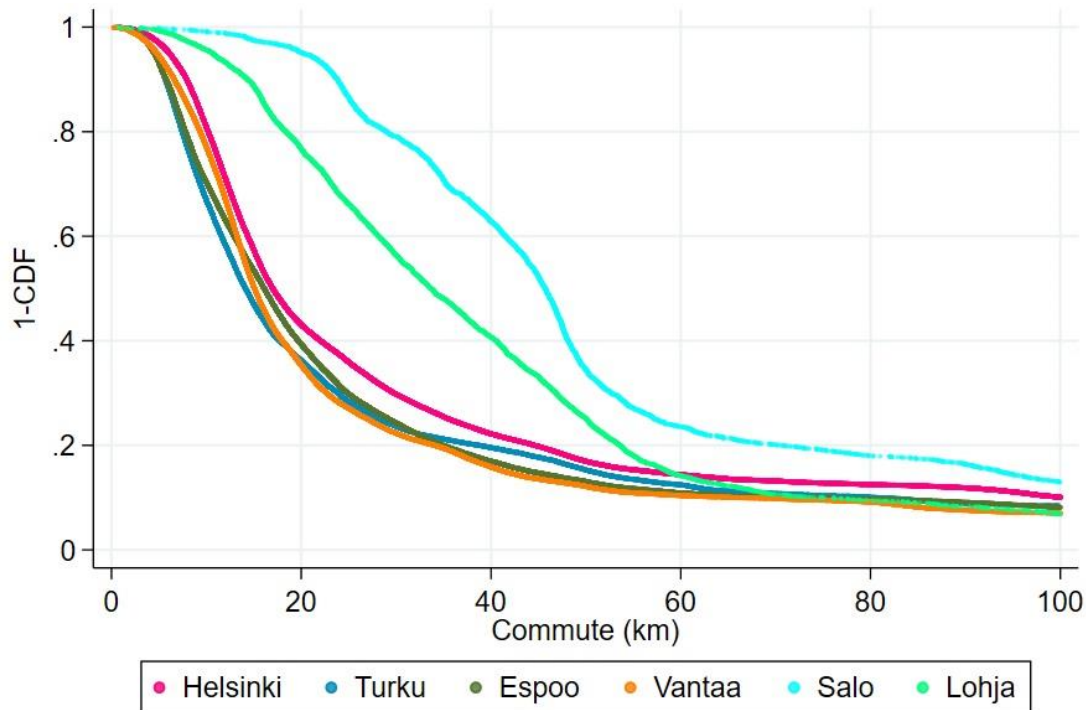
Kuvio 1. Helsinkiin suuntautuvan työmatkan kehitys vuosina 2005, 2010 ja 2015



Kuvio 2. Turkuun suuntautuvan työmatkan kehitys vuosina 2005, 2010 ja 2015

Kuviossa 3 on esitettyä samassa kuviossa aineiston päättymisvuoden, eli 2017 vastaavat jakaumat. Kuvion 3 perusteella suuriin kaupunkeihin pendelöi noin 20% otoksesta pidemmältä kuin 40 kilometrin päästä. 60 kilometriä tai pidemmältä

pendelöivien osuus on enää 15%. Tästä pidemmät työmatkat edustavat jakauman häntää, missä ei enää tapahdu kovinkaan suuria muutoksia. Salo ja Lohja ovat tyypillisempiä syrjäisempiä seutuja, mihin pendelöidään pidempien matkojen päästä. Salon ja Lohjankin tapauksessa 60 kilometriä näyttäisi olevan samanlainen kynnsarvo, minkä jälkeen pidemmät työmatkat edustavat vain häntää. 60 kilometriä vaikuttaisi luvussa 2 mainitun 100 kilometrin lisäksi olevan hyvä kaukopenelöintiä määrittävä kynnsarvo deskriptiivisen tarkastelun perusteella.



Kuvio 3. Muutamiin kuntiin kohdistuvien työmatkojen tiheysfunktioiden jakauma matkan mukaan

4.3 Deskriptiivinen tarkastelu pendelöinnin kestosta

Pendelöinnin kestosta tehtiin deskriptiivinen tarkastelu eräiden kiinnostuksen kohteena olevien muuttujien osalta. Tässä tarkastelussa kaukopenelöinti periodit ovat summattuna yhteen tarkastelujaksolta, ja sen jälkeen laskettu keskiarvo muutamalle kiinnostuksen kohteena käytetyille muuttujille. Taulukossa 2 näkyy näin lasketut kaukopenelöinnin kestot 60 kilometrin kynnsarvolla, ja taulukossa 3 näkyy vastaava 100 kilometrin kynnsarvolla.

Taulukosta 2 on havaittavissa, että tarkastelujaksolta sellaiset henkilöt, jotka kaukopenelöivät 60km kynnsarvolla, kaukopenelöivät keskimäärin 3 periodia, eli

kolme vuotta. Tulokset ovat saman kaltaisia kuin edellisissäkin malleissa.

Tutkijakoulutetut ja miehet korostuvat kaukopenelöivienkin joukossa. Yllättävä havainto on paikkakunnalta toiselle muuttajan ajassa lyhyt pendelöinnin kesto. Tämä voisi kuvastaa sitä, että muuttajat pendelöivät vanhaan työpaikkaansa sen aikaa, kunnes löytävät lähempää uutta asuinkuntaansa uuden. 100 kilometrin kynnyksellä havainnot ovat likimain samat kuin 60 kilometrinkin kynnyksellä.

Taulukko 2. 60km kaukopenelöijien pendelöinnin yhteenlaskettujen kestojen keskiarvot

60km	Keskiarvo	Keskihajonta	N
Kaikki kaukopenelöijät	3.00	2.32	42374
Henkilökohtaiset muuttajat			
Tutkijakoulutus	3.43	2.48	1174
Ylempi korkeakoulutus	3.19	2.36	8878
Alempikorkeakoulutus	2.96	2.25	7075
Muuttaja	2.69	2.09	8550
Työntekijä	2.84	2.25	9491
Pääkaupunkiseudulla asuva	2.63	2.03	14899
Mies	3.21	2.43	24789
Asuinpaikkakohtaiset muuttajat			
VT1	2.88	2.22	25238
Rantarata	2.87	2.21	24483
Päärata	2.61	2.03	3287

Taulukko 3. 100km kaukopenelöijien yhteenlaskettujen kestojen keskiarvot

100km	Keskiarvo	Keskihajonta	N
Kaikki kaukopenelöijät	2.79	2.19	28572
Henkilökohtaiset muuttajat			
Tutkijakoulutus	3.31	2.41	953
Ylempi korkeakoulutus	2.95	2.24	6277
Alempi korkeakoulutus	2.74	2.13	4897
Muuttaja	2.40	1.89	5228
Työntekijä	2.63	2.11	5857
Pääkaupunkiseudulla asuva	2.48	1.91	10988
Mies	2.99	2.30	17151
Asuinpaikkakohtaiset muuttajat			
VT1	2.72	2.12	18973
Rantarata	2.68	2.08	17774
Päärata	2.43	1.93	2188

5 Kaukopenelöintiä selittäviä tekijöitä

5.1 Probit- mallien tuloksia

Estimoidaan kaksi Probit- mallia, jossa toisessa on Helminen & Ristimäki (2007), sekä Ruotsin Trafik analyysin käyttämä 100 kilometrin kynnysarvo. Toisessa käytetään kuvion 1 ja 2 määrittämää 60 kilometrin kynnysarvoa. Tämän tarkastelun avulla on mahdollista saada selville, minkälaisia ihmisiä kaukopenelöijät ovat.

100 kilometrin kynnysarvoa käyttävän mallin tulokset ovat raportoituna taulukossa 4. Lähes kaikki muuttajat saavat tilastollisen merkitsevyyden, pois lukien ulkomaalaistausta, juokseva vuosi ja toisen asteen korkein tutkinto. Taulukon 4 mukaisesti kaukopenelöijien määrä tarkastelujaksolla ei ole juurikaan muuttunut. Henkilökohtaisista muuttujista mies- sukupuoli, sekä korkea koulutus lisäävät vahvasti kaukopenelöinnin todennäköisyyttä. Erityisesti tutkijakoulutuksen käyminen nostaa todennäköisyyttä sille, että henkilö on kaukopenelöijä. Toisaalta kiinnostava havainto on, että sosioekonomisella asemalla ei näyttäisi olevan vastaavan kaltaista ominaisuutta kaukopenelöinnin todennäköisyyden selittäjänä. Korkea koulutus ja ylempi toimihenkilö- status eivät mene pendelöinnin suhteen rinta rintaan, kuten olisi voinut ajatella. Asumismuodolla ei ole suurta merkitystä tarkastelussa. Yllättävää tässä aineistossa on, että vuokra-asuminen lisääisi pendelöinnin todennäköisyyttä, kun taas omistusasuminen vähentäisi sitä. Tämä saattaisi selittyä työpaikan lähelle muuttamisella: työpaikkaan ollaan valmiimpia sitoutumaan enemmän, mikäli omistaa talon läheltä työpaikkaa. Tämän toisaalta tulisi näkyä siinä, että eri kuntaan muuttamisen tulisi laskea todennäköisyyttä kaukopenelöinnille. Näin ei kuitenkaan ole. Eri kuntaan muuttamisen korkean kaukopenelöinnin todennäköisyyttä lisäävät kertoimen voisi ehkä selittää se, että muuttopäätöksen jälkeen pendelöidään vielä hetki vanhaan työpaikkaan, kunnes löydetään lähempää uusi.

Asuinpaikkaan liittyvistä muuttujista asuminen pääkaupunkiseudulla vähentää kaukopenelöinnin todennäköisyyttä merkittävästi. Liikenneyhteydet, kuten Rantarata ja VT1 lisäävät kaukopenelöinnin todennäköisyyttä. VT1 melko suurella kertoimella. Pääradalla ei ole tilastollista merkitsevyyttä, mutta pääradasta tarkastelualueella on vain pieni osa.

Taulukko 4. Yli 100 kilometrin työmatkoja selittävän probit- mallin tulokset

	Coef.	Std. Err.	z	P> z
Juokseva vuosi	0.001	0.000	1.76	0.079
Henkilökohtaiset muuttujat				
Ikä	-0.003	0.000	-21.2	0
Mies	0.280	0.003	103.1	0
Työttömyyskuukausia vuoden aikana	0.057	0.002	29.73	0
Aviossa	-0.047	0.003	-16.19	0
Omistusasuja	-0.033	0.006	-5.25	0
Vuokra-asuja	0.045	0.007	6.72	0
Käytössä olevat tulot tuhansina	0.004	0.000	54.92	0
Alle 7- vuotiaiden lasten lukumäärä	-0.042	0.002	-20.4	0
Tulojen muutos t-1 ja t välillä	-0.002	0.000	-5.3	0
Eri kuntaan muuttaja	0.273	0.005	53.93	0
Ylempi toimihenkilö	0.164	0.005	33.35	0
Alempi toimihenkilö	0.045	0.005	9.36	0
Työntekijä	0.047	0.005	9.2	0
Ulkomaalaistausta	0.002	0.006	0.37	0.715
Korkein tutkinto toinen aste	0.006	0.004	1.35	0.178
Korkein tutkinto erikoisammattikoulu	-0.052	0.014	-3.65	0
Korkein tutkinto alin korkea-aste	0.069	0.005	12.97	0
Korkein tutkinto alempi korkeakoulu	0.077	0.005	14.6	0
Korkein tutkinto ylempi korkeakoulu	0.113	0.005	20.93	0
Korkein tutkinto tutkijakoulutus	0.406	0.008	51.6	0
Asuinpaikkakohtaiset muuttujat				
Asuminen pääkaupunkiseudulla	-0.467	0.003	-146.38	0
Rantarata asuinkunnassa	0.043	0.004	9.93	0
Päärata asuinkunnassa	-0.037	0.005	-7.51	0
VT1 asuinkunnassa	0.170	0.004	39.22	0
_cons	-2.342	0.011	-215.68	0
N	7027664			
LR chi2(25)	59395			
Prob>chi2	0.0000			
Pseudo R2	0.052			

Taulukossa 5 on kuvattuna vastaavanlainen probit- malli estimoituna 60 kilometrin kynnyksarvolla. Käytettäessä 60 kilometrin kynnyksarvoa tilanne muuttuu joiltain osin. VT1 selittää pendelöintiä positiivisella kertoimella yli sadan kilometrin kynnyksarvolla, mutta yli 60 kilometrin arvolla kerroin on negatiivinen. On jokseenkin odotettavaa, että VT1 auttaa juuri pitkän matkan pendelöintiin. Korkeakouluttujen osalta tulos on sama kuin edellisessä sadan kilometrin kynnyksarvon mallissa, mutta sosioekonominen asema näyttäisi 60 kilometrin tapauksessa menevän enemmän korkean koulutuksen kanssa käsi kädessä, sillä ylempi toimihenkilyys saa korkeamman kertoimen verrattuna 100

kilometrin malliin. Rantarata saa huomattavasti suuremman kertoimen 60 kilometrin mallissa kuin 100 kilometrin mallissa. Voisi ajatella, että ihmisiä siis ei pendelöi Rantarataa pisin Turun ja Helsingin välillä, vaan lyhyempiä matkoja joko Helsinkiin tai Turkuun näiden välisistä kunnista.

Näiden kahden mallin perusteella kaukopendelöijän profiiliksi voisi kuvailla pääkaupunkiseudun ulkopuolella asuva korkeakoulutettu mies, kuka on juuri muuttanut alueelle. Hieman yllättävästi pidemällä matkalla VT1 kotikunnassa nostaa pendelöintitodennäköisyyttä, ja lyhyemmällä matkalla Rantaradan oleminen kotikunnassa. Tämä motivoi hyvin tässä tutkielmassa esitettäviin jatkotutkimuksiin tekijöistä, mitkä laukaisevat kaukopendelöinnin.

Taulukko 5. Yli 60 kilometrin työmatkoja selittävän probit- mallin tulokset

	Coef.	Std. Err.	z	P> z
Juokseva vuosi	0.003	0.000	9.76	0
Henkilökohtaiset muuttujat				
Ikä	-0.002	0.000	-19.64	0
Mies	0.290	0.002	128.17	0
Työttömyyskuukausia vuoden aikana	0.059	0.002	35.88	0
Aviossa	-0.055	0.002	-22.7	0
Omistusasuja	0.001	0.005	0.14	0.888
Vuokra-asuja	0.053	0.006	9.12	0
Käytössä olevat tulot tuhansina	0.004	0.000	63.93	0
Alle 7- vuotiaiden lasten lukumäärä	-0.032	0.002	-18.56	0
Tulojen muutos t-1 ja t välillä	-0.002	0.000	-5.49	0
Eri kuntaan muuttaja	0.330	0.004	79.03	0
Ylempi toimihenkilö	0.234	0.004	55.28	0
Alempi toimihenkilö	0.110	0.004	26.71	0
Työntekijä	0.136	0.004	31.77	0
Ulkomaalaistausta	0.016	0.005	3.17	0.002
Korkein tutkinto toinen aste	-0.009	0.004	-2.55	0.011
Korkein tutkinto erikoisammattikoulu	-0.031	0.011	-2.74	0.006
Korkein tutkinto alin korkea-aste	0.065	0.004	14.97	0
Korkein tutkinto alempi korkeakoulu	0.077	0.004	17.47	0
Korkein tutkinto ylempi korkeakoulu	0.125	0.005	27.82	0
Korkein tutkinto tutkijakoulutus	0.361	0.007	52.07	0
Asuinpaikkakohtaiset muuttujat				
Asuminen pääkaupunkiseudulla	-0.543	0.003	-196.97	0
Rantarata asuinkunnassa	0.196	0.004	54.84	0
Päärata asuinkunnassa	-0.135	0.004	-32.92	0
VT1 asuinkunnassa	-0.103	0.004	-28.73	0
_cons	-2.154	0.009	-235.23	0
N	7027664			
LR chi2(25)	108076.1			
Prob>chi2	0.0000			
Pseudo R2	0.062			

5.2 Ordered probit- mallin tuloksia

Aikaisemmissa tutkimuksissa nähdyn yksittäisen kaukopenelöinnin määrittävän kynnysarvon sijaan on ordered probit- malliin mahdollista ottaa useita eri kynnysarvoja työmatkoja selittävien tekijöiden mallintamiseen. Tässä tutkielmassa valitut arvot ovat mielivaltaisesti valittuja. Jaotteleamalla alle sadan kilometrin kaukopenelöinnin kynnysarvon jäävät matkat järjestyksellisiin suureisiin, on mahdollista tehdä tarkempaa

analyysiä mitä aikaisemmassa kirjallisuudessa on tehty. Tässä tutkielmassa käytetään seuraavanlaista jaottelua ordered probit- mallissa:

- Lyhyt työmatka väliltä (0,15] kilometriä
- Keskimääräinen työmatka väliltä [15,30) kilometriä
- Alle 40 minuuttia kestävä automatka, väli [30,60) kilometriä
- Alle tunnin kestävä automatka, väli [60,100) kilometriä
- Yli tunnin kestävä kaukopenelöinti, väli [100,300) kilometriä

Selittävät muuttujat ovat samat kuin edellisen alaluvun probit- mallissa. Taulukko 6 esittää aineiston jakautumisen valitun jaottelin määritteleville väleille. Aineistoa on riittävästi jokaiselle välille mallin ajamiseksi onnistuneesti.

Taulukko 6. Aineiston frekvenssit valituilla työmatkojen kynnysväleillä

Kynnysvälit	Frekvenss
(0,15]	5167672
[15,30)	1109261
[30,60)	561256
[60,100)	77390
[100,300)	112085
Yhteensä	7027664

Taulukossa 7 on esitetty ordered probit- mallin tuloksia. Kaikki muuttujat ovat tässä tarkastelussa tilastollisesti merkitseviä. Tarkastelussa havaitaan monia eroja edellisiin probit- mallin antamiin tuloksiin nähden. Tässä mallissa luonnollisesti sosioekonominen aseman kerroin nousee suureksi. Pitää olla töissä, jotta voi pendelöidä. Työssä käynti siis luonnollisesti selittää työmatkoja. Kiinnostavia havaintoja on edelleen miesten suurempi pendelöinti, pääkaupunkiseudun suuri pendelöintitodennäköisyyttä vähentävä vaikutus ja eri kuntaan muuttamisen pysyminen suurella kertoimella. Tässä tarkastelussa VT1 näyttäisi vähentävän pendelöintitodennäköisyyttä, mikä voisi selittyä pääkaupunkiseudun erittäin alhaisella pendelöintitodennäköisyydellä. Asumismuotojen kertoimet ovat tässä mallissa odotetumpia kuin edellisessä malleissa omistusasumisen lisätessä pendelöintiä ja vuokra-asumisen vähentäessä sitä.

Taulukko 7. Ordered probit- mallin tuloksia

	Coef.	Std. Err.	z	P> z
Juokseva vuosi	0.005	0.000	35.66	0
Henkilökohtaiset muuttujat				
Ikä	-0.004	0.000	-72.48	0
Mies	0.230	0.001	206.12	0
Työttömyyskuukausia vuoden aikana	0.031	0.001	32.1	0
Aviossa	0.005	0.001	4.56	0
Omistusasuja	0.036	0.003	14.06	0
Vuokra-asuja	-0.098	0.003	-34.77	0
Käytössä olevat tulot tuhansina	0.007	0.000	193.33	0
Alle 7- vuotiaiden lasten lukumäärä	0.027	0.001	32.61	0
Tulojen muutos t-1 ja t välillä	-0.002	0.000	-8.52	0
Eri kuntaan muuttaja	0.374	0.002	160.09	0
Ylempi toimihenkilö	0.678	0.002	294.3	0
Alempi toimihenkilö	0.621	0.002	283.77	0
Työntekijä	0.593	0.002	259.56	0
Ulkomaalaistausta	0.012	0.003	4.82	0
Korkein tutkinto toinen aste	-0.029	0.002	-17.13	0
Korkein tutkinto erikoisammattikoulu	0.062	0.005	12.07	0
Korkein tutkinto alin korkea-aste	0.095	0.002	45.59	0
Korkein tutkinto alempi korkeakoulu	0.080	0.002	37.33	0
Korkein tutkinto ylempi korkeakoulu	0.049	0.002	21.61	0
Korkein tutkinto tutkijakoulutus	0.065	0.004	15.54	0
Asuinpaikkakohtaiset muuttujat				
Asuminen pääkaupunkiseudulla	-0.696	0.001	-493.43	0
Rantarata asuinkunnassa	-0.128	0.002	-66.7	0
Päärata asuinkunnassa	0.040	0.002	22.33	0
VT1 asuinkunnassa	-0.265	0.002	-141.08	0
/cut1	0.983	0.005		
/cut2	1.693	0.005		
/cut3	2.443	0.005		
/cut4	2.662	0.005		
N	7027664			
LR chi2(25)	1060603			
Prob>chi2	0			
Pseudo R2	0.090			

Eroina lopputuloksissa ordered probit- sekä probit- mallin välillä yleisesti sitä, että ordered probitissa on määrällisesti enemmän tilastollisesti merkitseviä tekijöitä. Tilastollisen merkitsevyyden ulkopuolelle jää ordered probitissa vain juokseva vuosi, alle 7 vuotiaiden lasten lukumäärä, ulkomaalaistausta sekä erikoisammattikoulun käyminen.

Ordered probit- mallissa henkilökohtaisista muuttujista miessukupuoli on edelleen suurella positiivisella kertoimella selittävänä tekijänä aivan kuin probit- malleissakin. Ikä, avioliitossa oleminen, matala koulutus, vuokra-asuminen sekä tulojen muutos laskee henkilön pendelöinti todennäköisyyttä ordered probit- mallissa.

Pendelöintitodennäköisyyttä nostavina henkilökohtaisina muuttujina nähdään toista astetta korkeampi koulutus probit- mallin tapaisesti nostamassa koulutusaste asteelta henkilön pendelöintitodennäköisyyttä suuremmaksi. Samoin muuttaminen eri kuntaan, omistusasuminen sekä korkeammat tulot lisäävät todennäköisyyttä pendelöinnille.

Kuntakohtaisista tekijöitä probit- mallien kanssa yhtenevä löydös on metropolialueella asumisen suuri vähentävä vaikutus pendelöintitodennäköisyyteen. Yllättävää on junaratojen henkilöpaikkojen sekä VT1 sijoittumisen asuinkuntaan pendelöintitodennäköisyyttä tilastollisesti merkitsevästi vähentävä vaikutus.

Taulukko 8 on tiivistelmä näiden kolmen mallin löydöksistä. Taulukkoon on riveille laitettu selittävät muuttujat, ja sarakkeisiin on eritelty, onko käytetty 60 kilometrin vai 100 kilometrin kynnsarvon probit- mallia vai ordered probittia.

Taulukko 8. Tiivistelmä probit- sekä ordered probit- mallien tuloksista

	60km	100km	Ordered probit
Juokseva vuosi	—	0	+
Henkilökohtaiset muuttajat			
Ikä	—	—	—
Mies	+	+	+
Työttömyyskuukausia vuoden aikana	+	+	+
Aviossa	—	—	+
Omistusasuja	0	—	+
Vuokra-asuja	+	+	—
Käytössä olevat tulot tuhansina	+	+	+
Alle 7- vuotiaiden lasten lukumäärä	—	—	+
Tulojen muutos t-1 ja t välillä	—	—	—
Eri kuntaan muuttaja	+	+	+
Ylempi toimihenkilö	+	+	+
Alempi toimihenkilö	+	+	+
Työntekijä	+	+	+
Ulkomaalaistausta	+	0	+
Korkein tutkinto toinen aste	—	0	—
Korkein tutkinto erikoisammattikoulu	—	—	+
Korkein tutkinto alin korkea-aste	+	+	+
Korkein tutkinto alempi korkeakoulu	+	+	+
Korkein tutkinto ylempi korkeakoulu	+	+	+
Korkein tutkinto tutkijakoulutus	+	+	+
Asuinpaikkakohtaiset muuttajat			
Asuminen metropolissa	—	—	—
Rantarata asuinkunnassa	+	+	—
Päärata asuinkunnassa	—	—	+
VT1 asuinkunnassa	—	+	—

0=ei tilastollista merkitsevyyttä p=0.05- merkitsevyytasolla.

+ = positiivinen tilastollisesti merkitsevä kerroin.

— = negatiivinen tilastollisesti merkitsevä kerroin.

Tämän perusteella voidaan profiloida todennäköinen kaukopenelöijä aviottomaksi, juuri eri paikkakunnalle muuttaneeksi korkeakoulutetuksi mieheksi, kuka ei asu pääkaupunkiseudulla. Todennäköisellä penelöijällä on myös työttömyyskuukausia takana, sekä korkeat tulot.

Aikaisemmasta tutkimuksesta penelöinnistä on löytynyt samaa. Muun muassa Marion & Hornerin (2007), Alasalmi ym. (2020) sekä Limtanakool ym. (2006) löydökset sukupuolen ja koulutuksen penelöinnin todennäköisyyttä nostavalle vaikutukselle löytyi myös jokaisesta kolmesta tämän tutkielman mallista. Alasalmen ym. (2020) löytämä pääkaupunkiseudun erityisen alhainen penelöinti tuli näissä malleissa esiin.

Myös Champion ym. (2009) löytämä muuttaja- statuksen pendelöintiä lisäävä vaikutus löytyi näistä malleista. Voisi myös ajatella, että mahdollisuus etätyöhön kasvaa koulutustason myötä, jolloin työpaikka on mahdollista ottaa kauempaa vastaan. Tämä selittäisi Helminen & Ristimäen (2007) löydöstä etätyön lisäävästä vaikutuksesta.

Alasalmi ym. (2020) vastaisesti tässä aineistossa matala koulutus ei ennustanut kovinkaan merkittävästi liikkuvuutta verrattuna korkeakoulutettuihin. Toisaalta, tulokset voisivat muuttua, mikäli pendelöinnin tämän tutkielman määrittelystä poistaisimme tietoverkkopendelöinnin kokonaan. Nämä tulokset eivät siis poissulje sitä mahdollisuutta, että matala koulutus voisi selittää fyysistä pendelöintiä. Geslin & Ravaletin (2016) löytämän ankkurointivaikutuksen puolesta voisi puhua molemmissa kaukopenelöintimalleissa löytynyt avioliiton tilastollisesti merkitsevät miinusmerkkiset kertoimet ja ordered probit- mallin hyvin pieni positiivinen kerroin. Avioliitto voisi olla eräs tekijä juurtumisessa tiettyyn asuinkuntaan.

6 Pendelöinnin laukaisevia tekijöitä

6.1 Muutokset malliin

Edellisessä luvussa olleiden mallien tavoitteena oli profiloida, millainen henkilö pendelöijä on. Näissä malleissa oli probit- malleihin koodattu joko 1 = pendelöijä, tai 0 = ei ole pendelöijä, erinäisin kynnsarvoin. Tämän luvun mallien tavoitteena on selvittää, minkä tyyppisiä henkilöt ovat, jotka aloittavat tai lopettavat pendelöinnin.

Pendelöinnin aloitus voidaan probit- mallien tapauksessa määrittää myös kynnsarvolla. Tässä tarkastelussa kynnsarvot ovat vastaavat 100 kilometriä, sekä 60 kilometriä, mitä edellisessäkin tarkastelussa. Ordered probit- mallissa käytetään myös samoja kynnsarvoja työmatkalle kuin edellisessä luvussa. Valitut muuttujat ovat myös samat, jotta tulokset pysyvät vertailukelpoisina keskenään.

Erona edellisen luvun malleihin kuitenkin on aineiston jaottelu pendelöinnin sijaan jaottelu työmatkan muutosten mukaan viime vuoteen nähden. Työmatkan muutoksesta voidaan havaita, milloin aineiston henkilö aloittaa kaukopendelöinnin, ja milloin lopettaa. Mikäli työmatkan muutos edellisen ja nykyisen periodin välillä on enemmän kuin valittu kynnsarvo, henkilö koodataan arvolla 1. Muutoin 0.

Toinen kiinnostuksen kohde on pendelöinnin lopettaminen. Mallissa kaikki asiat ovat samoin kuin edellisessä kappaleessa on selitettynä, mutta työmatkan kasvamisen sijaan tarkastellaan työmatkan vähentymistä. Henkilö koodataan tällöin arvolla 1, jos työmatka lyhenee valitun kynnsarvon verran. Muutoin henkilö koodataan nolllalla. Kynnsarvot ovat itseisarvoltaan samat kuin edellisessä kohdassa, mutta arvoltaan miinusmerkkiset.

Tämäkään mallin avulla ei ole mahdollista tehdä kausaalipäätelyä mitkä tekijät ovat aiheuttaneet pendelöinnin aloittamisen. Mallin tuloksia voi tulkita edellisten mallien tapaan profilointina ja vastaa kysymykseen ”Millaiset ihmiset aloittavat tai lopettavat pendelöinnin?”.

6.2 Pendelöinnin aloitusta selittävien probit- mallien tuloksia

Taulukossa 9 on probit- mallin tuloksia 100 kilometrin kynnsarvolla. Muuttujat saavat hyvin selitysarvoa, vain Päärata on tilastollisesti merkitsemätön. Eri kuntaan muuttaminen näyttäisi olevan eräs kaukopendelöinnin aloitusta selittävä tekijä. Tämä

voisi vahvistaa edellisessä luvussa tehtyä ajatusta siitä, että eri kunnasta käydään vielä hetken aikaa vanhassa työpaikassa pitkänkin matkan päässä töissä. Korkeakoulutus näyttäisi myös olevan yhteydessä pendelöinnin aloittamiseen, kuten edellisissä malleissa pendelöintiin yleisesti. Työpaikan saanti työttömyysjakson jälkeen kauempaakin voisi olla melko luonnollinen selitys kaukopenelöinnin aloittamiselle. Tulojen muutoksella ei ole tilastollista merkitsevyyttä, mikä on jossain määrin yllättävä löydös ajatellen siten, että parempi palkkaisen työpaikan saaminen kauempaa voisi liipaista henkilön kaukopenelöintiin.

Asuinpaikkakohtaisissa muuttujissa edellistä mallia mukailien, asuminen metropolissa toimii kaukopenelöinnin aloittamistodennäköisyyttä vähentävänä tekijänä. Rantaradan henkilöliikennepaikan oleminen asuinkunnassa lisää kaukopenelöinnin aloittamisen todennäköisyyttä tilastollisesti merkitsevästi, kuten myös VT1 oleminen kotikunnassa.

Taulukko 9. Probit- mallin tulokset työmatkan periodittaisen muutoksen ollessa yli 100 kilometriä

	Coef.	Std. Err.	z	P> z
Juokseva vuosi	0.001	0.001	1.05	0.295
Henkilökohtaiset muuttujat				
Ikä	-0.010	0.000	-40.93	0
Mies	0.188	0.005	40.16	0
Työttömyyskuukausia vuoden aikana	0.104	0.002	43.98	0
Aviossa	-0.027	0.005	-5.38	0
Omistusasuja	-0.039	0.011	-3.64	0
Vuokra-asuja	0.070	0.011	6.18	0
Käytössä olevat tulot tuhansina	0.001	0.000	10.29	0
Alle 7- vuotiaiden lasten lukumäärä	-0.056	0.003	-15.94	0
Tulojen muutos t-1 ja t välillä	-0.002	0.001	-2.81	0.005
Eri kuntaan muuttaja	0.531	0.007	81.14	0
Ylempi toimihenkilö	0.067	0.008	8.03	0
Alempi toimihenkilö	-0.059	0.008	-7.29	0
Työntekijä	-0.115	0.009	-13.07	0
Ulkomaalaistausta	-0.056	0.011	-5.18	0
Korkein tutkinto toinen aste	0.019	0.008	2.38	0.017
Korkein tutkinto erikoisammattikoulu	0.052	0.024	2.2	0.028
Korkein tutkinto alin korkea-aste	0.068	0.010	6.96	0
Korkein tutkinto alempi korkeakoulu	0.072	0.009	7.69	0
Korkein tutkinto ylempi korkeakoulu	0.132	0.010	13.82	0
Korkein tutkinto tutkijakoulutus	0.329	0.014	23.17	0
Asuinpaikkakohtaiset muuttujat				
Asuminen pääkaupunkiseudulla	-0.310	0.005	-57.08	0
Rantarata asuinkunnassa	0.050	0.008	6.4	0
Päärata asuinkunnassa	0.015	0.009	1.73	0.083
VT1 asuinkunnassa	0.155	0.008	19.74	0
_cons	-2.445	0.019	-131.5	0
N	7027664			
LR chi2(25)	19522			
Prob>chi2	0.0000			
Pseudo R2	0.056			

Taulukossa 10 on probit- mallin tuloksia 60 kilometrin kynnsarvolla. Mallin tuloksissa päälinja on sama, mitä 100 kilometrin kynnsarvon mallissa. Tämänkin mallin muuttujat saavat hyvin tilastollista merkitsevyyttä. Eri kuntaan muuttajan kerroin näyttäisi nousevan lyhyemmällä työmatkan muutoksella. Tutkijakoulutuksella on edelleen suuri kerroin.

Taulukko 10. Probit- mallin tulokset työmatkan periodittaisen muutoksen kynnyksarvon ollessa 60 kilometriä

	Coef.	Std. Err.	z	P> z
Juokseva vuosi	0.002	0.001	2.9	0.004
Henkilökohtaiset muuttujat				
Ikä	-0.009	0.000	-45.3	0
Mies	0.182	0.004	45.6	0
Työttömyyskuukausia vuoden aikana	0.111	0.002	54	0
Aviossa	-0.028	0.004	-6.52	0
Omistusasuja	-0.030	0.009	-3.24	0.001
Vuokra-asuja	0.063	0.010	6.38	0
Käytössä olevat tulot tuhansina	0.002	0.000	12.56	0
Alle 7- vuotiaiden lasten lukumäärä	-0.052	0.003	-17.35	0
Tulojen muutos t-1 ja t välillä	-0.002	0.001	-2.93	0.003
Eri kuntaan muuttaja	0.616	0.005	112.63	0
Ylempi toimihenkilö	0.075	0.007	10.35	0
Alempi toimihenkilö	-0.050	0.007	-7.18	0
Työntekijä	-0.081	0.007	-10.9	0
Ulkomaalaistausta	-0.045	0.009	-4.97	0
Korkein tutkinto toinen aste	0.002	0.007	0.25	0.801
Korkein tutkinto erikoisammattikoulu	0.038	0.020	1.88	0.06
Korkein tutkinto alin korkea-aste	0.051	0.008	6.26	0
Korkein tutkinto alempi korkeakoulu	0.058	0.008	7.36	0
Korkein tutkinto ylempi korkeakoulu	0.119	0.008	14.79	0
Korkein tutkinto tutkijakoulutus	0.275	0.013	21.77	0
Asuinpaikkakohtaiset muuttujat				
Asuminen pääkaupunkiseudulla	-0.332	0.005	-70.32	0
Rantarata asuinkunnassa	0.132	0.007	19.93	0
Päärata asuinkunnassa	-0.031	0.007	-4.17	0
VT1 asuinkunnassa	0.031	0.007	4.7	0
_cons	-2.307	0.016	-145.32	0
N	7027664			
LR chi2(25)	29866			
Prob>chi2	0.0000			
Pseudo R2	0.061			

6.3 Pendelöinnin aloitusta selittävän ordered probit- mallin tuloksia

Taulukon 11 mukaisilla kynnyksväleillä estimoitiin pendelöinnin muutosta. Nämä kynnyksvälit kuvastavat, miten pendelöinnin muutos jakautuu aineistoon. Aineistosta on tässä kohtaa poistettu ne, jotka lopettavat pendelöinnin, tai eivät muuta pendelöintiään. Pendelöinnin lopettaneisiin palataan myöhemmin tässä tutkielmassa.

Aineisto jakautuu ennustetusti painottuen lyhyisiin pendelöintimatkojen muutoksiin. Kiinnostava havainto aineistosta on, että 100-300 kilometrin muutoksessa on noin

kaksinkertaisesti aineistoa verrattuna 60-100 kilometrin muutokseen. Estimoinnin kannalta oleellista on, että jokaisella kynnysvälillä aineistoa on riittävästi.

Taulukko 11. Ordered probit- mallin frekvenssit valituille kynnysväleille pendelöinnin aloittajissa

Kynnysvälit	Frekvenssi
(0,15]	1229667
[15,30)	74918
[30,60)	44847
[60,100)	13216
[100,300)	26603
Yhteensä	1389251

Taulukossa 12 on kuvattuna edellä mainitulla aineistolla ajetun ordered probit- mallin tuloksia. Selittävät muuttujat ovat samat kuin edellisissäkin malleissa. Tässä mallissa huomionarvoista on se, että tilastollista merkitsevyyttä löytyy erittäin monesta selittävästä muuttujasta. Mallin tulkinta on sama kuin edellisissä probit- malleissa. Malli kykenee tuomaan esiin niitä tekijöitä, mitkä lisäävät pendelöinnin aloittamisen todennäköisyyttä. Kausaalipäätelyä mallin pohjalta on mahdotonta tehdä.

Henkilökohtaisista muuttujista eri kuntaan muuttaja saa tässäkin tarkastelussa erittäin suuren kertoimen. Myös tutkijakoulutuksen saaneet saavat suuren kertoimen. Muuttajat ovat siis pendelöinnin kannalta hyvin kiinnostava ryhmä. Asumispaikkakohtaisissa muuttujissa edeltävien mallien mukaisesti pääkaupunkiseudulla asuminen vähentää reilusti pendelöinnin aloittamistodennäköisyyttä.

Taulukko 12. Ordered probit- mallin tuloksia pendelöinnin aloittavista

	Coef.	Std. Err.	z	P> z
Juokseva vuosi	0.002	0.000	3.92	0
Henkilökohtaiset muuttujat				
Ikä	-0.006	0.000	-37.08	0
Mies	0.164	0.003	52.35	0
Työttömyyskuukausia vuoden aikana	0.101	0.002	56.64	0
Aviossa	-0.004	0.003	-1.05	0.295
Omistusasuja	0.029	0.007	4.03	0
Vuokra-asuja	0.058	0.008	7.67	0
Käytössä olevat tulot tuhansina	0.001	0.000	4.86	0
Alle 7- vuotiaiden lasten lukumäärä	-0.044	0.002	-19.48	0
Tulojen muutos t-1 ja t välillä	-0.001	0.001	-2.14	0.033
Eri kuntaan muuttaja	0.753	0.004	205.3	0
Ylempi toimihenkilö	-0.043	0.006	-6.94	0
Alempi toimihenkilö	-0.116	0.006	-19.71	0
Työntekijä	-0.110	0.006	-17.9	0
Ulkomaalaistausta	-0.038	0.007	-5.56	0
Korkein tutkinto toinen aste	-0.022	0.005	-4.49	0
Korkein tutkinto erikoisammattikoulu	0.068	0.015	4.7	0
Korkein tutkinto alin korkea-aste	0.031	0.006	4.8	0
Korkein tutkinto alempi korkeakoulu	0.019	0.006	3.17	0.002
Korkein tutkinto ylempi korkeakoulu	0.092	0.006	14.18	0
Korkein tutkinto tutkijakoulutus	0.205	0.012	17.22	0
Asuinpaikkakohtaiset muuttujat				
Asuminen pääkaupunkiseudulla	-0.586	0.004	-149.31	0
Rantarata asuinkunnassa	0.013	0.005	2.37	0.018
Päärata asuinkunnassa	-0.104	0.005	-19.89	0
VT1 asuinkunnassa	-0.104	0.005	-20.01	0
/cut1	0.852	0.012		
/cut2	1.241	0.012		
/cut3	1.615	0.013		
/cut4	1.787	0.013		
N	1389251			
LR chi2(25)	111539.5			
Prob>chi2	0.0000			
Pseudo R2	0.081			

6.4 Pendelöinnin lopettamista selittävien probit- mallien tuloksia

Pendelöinnin lopettamista selittävät mallit ovat rakennettu samalla tavalla kuin pendelöinnin aloittamista selittävät mallit. Tehdään tarkastelu kahdella probit- mallilla sekä ordered probit- mallilla.

Probit- mallin kynnyksarvoina on tässä tarkastelussa työmatkan väheneminen 60 kilometrillä, sekä 100 kilometrillä. Tämä kuvastaa henkilön kaukopendelöinnin

oletettua lopettamista, tai ainakin merkittävää vähentämistä. Selittävät muuttajat ovat samat kuin edellisissäkin malleissa. Taulukossa 13 on kuvattuna probit- mallin tuloksia. Tässä selitettävä muuttaja on koodattu 1, mikäli työmatka edeltäneen ja tarkasteluvuoden välillä on vähentynyt yli 100 kilometriä. Muutoin nolla.

Henkilökohtaisista muuttajista ehkä hieman yllättävästi korkeakoulutus, sekä erityisesti tutkijakoulutus, näyttäisi nostavan myös pendelöinnin lopettamisen todennäköisyyttä 100 kilometrin kynnyksellä. Selitys tälle löydökselle on todennäköisesti hyvin triviaali: Pendelöinti pitää aloittaa, jotta sen voi lopettaa.

Pelkästään itse kaukopendelöintiä selittäneet mallit toivat esiin sen, että kaukopendelöijä ovat korkeakoulutetut, ja erityisesti tutkijakoulutuksen saaneet. Pendelöinnin aloittavat kuuluvat samoihin ryhmiin. Jos nämä ryhmät korostuvat kaukopendelöinnin aloittaneissa, sekä kaukopendelöivissä, näiden ryhmien voisi olettaa korostuvan myös pendelöinnin lopettaneissa. Vastaava päättely voitaneen tehdä myös mies- sukupuolen korostumisesta kaukopendelöinnin lopettaneissa. Korkeakoulutetut, erityisesti tutkijakoulutuksen käyneet näyttäisivät siis olevan hyvin dynaaminen ryhmä.

Muuttaja- ryhmä nousee tässäkin tarkastelussa mielenkiintoisimmaksi. Edellisen kappaleen selitys ei välttämättä ole pätevä muuttajien tapauksessa, sillä muuttamisen ajanhetki on tietty vuosi, jolloin kotikunta on muuttunut. Muuttaminen siis toimii molempiin suuntiin: joko pendelöinnin laukaisevana tekijänä tai pendelöinnin lopettavana tekijänä. Selitys voisi olla työpaikan löytämisen ja muuttamisen tapahtumajärjestyksessä. Toiset muuttavat työpaikan perässä ja toiset muuttavat ensin ja sen jälkeen etsivät läheltä työpaikan. Paikkakuntakohtaisissa muuttajissa korostuu tässäkin tarkastelussa erityisesti pääkaupunkiseudun asukkaiden hyvin matala liikkuvuus.

Taulukko 13. Probit- mallin tuloksia työmatkan periodittaisen muutoksen ollessa -100km

	Coef.	Std. Err.	z	P> z
Juokseva vuosi	-0.003	0.001	-4.57	0
Henkilökohtaiset muuttujat				
Ikä	-0.012	0.000	-53.18	0
Mies	0.171	0.004	38.27	0
Työttömyyskuukausia vuoden aikana	0.100	0.002	43.19	0
Aviossa	-0.026	0.005	-5.37	0
Omistusasuja	-0.069	0.010	-6.91	0
Vuokra-asuja	0.079	0.011	7.51	0
Käytössä olevat tulot tuhansina	0.000	0.000	0.63	0.532
Alle 7- vuotiaiden lasten lukumäärä	-0.049	0.003	-14.79	0
Tulojen muutos t-1 ja t välillä	-0.001	0.001	-0.76	0.447
Eri kuntaan muuttaja	0.371	0.007	53.38	0
Ylempi toimihenkilö	-0.078	0.008	-10.11	0
Alempi toimihenkilö	-0.147	0.007	-19.98	0
Työntekijä	-0.209	0.008	-26.05	0
Ulkomaalaistausta	-0.048	0.010	-4.87	0
Korkein tutkinto toinen aste	0.028	0.008	3.63	0
Korkein tutkinto erikoisammattikoulu	0.045	0.024	1.87	0.061
Korkein tutkinto alin korkea-aste	0.093	0.010	9.75	0
Korkein tutkinto alempi korkeakoulu	0.130	0.009	14.57	0
Korkein tutkinto ylempi korkeakoulu	0.210	0.009	22.84	0
Korkein tutkinto tutkijakoulutus	0.393	0.014	28.04	0
Asuinpaikkakohtaiset muuttujat				
Asuminen pääkaupunkiseudulla	-0.236	0.005	-44.76	0
Rantarata asuinkunnassa	0.027	0.008	3.38	0.001
Päärata asuinkunnassa	0.017	0.008	2.11	0.035
VT1 asuinkunnassa	0.156	0.008	19.69	0
_cons	-2.150	0.017	-122.95	0
N	7027664			
LR chi2(25)	17320			
Prob>chi2	0.0000			
Pseudo R2	0.046			

Taulukossa 14 on kuvattuna vastaavan mallin tuloksia -60- kilometrin kynnyksarvolla. Suuria eroja ei mallien välille näytä tulevan.

Taulukko 14. Probit- mallin tuloksia työmatkan periodittaisen muutoksen ollessa -60 kilometriä

	Coef.	Std. Err.	z	P> z
Juokseva vuosi	-0.002	0.001	-3.89	0
Henkilökohtaiset muuttujat				
Ikä	-0.012	0.000	-60.97	0
Mies	0.170	0.004	43.89	0
Työttömyyskuukausia vuoden aikana	0.104	0.002	50.94	0
Aviossa	-0.029	0.004	-7.03	0
Omistusasuja	-0.065	0.009	-7.41	0
Vuokra-asuja	0.079	0.009	8.56	0
Käytössä olevat tulot tuhansina	0.000	0.000	0.03	0.975
Alle 7- vuotiaiden lasten lukumäärä	-0.042	0.003	-14.74	0
Tulojen muutos t-1 ja t välillä	0.000	0.001	0.01	0.989
Eri kuntaan muuttaja	0.419	0.006	70.54	0
Ylempi toimihenkilö	-0.093	0.007	-13.95	0
Alempi toimihenkilö	-0.169	0.006	-26.65	0
Työntekijä	-0.212	0.007	-30.96	0
Ulkomaalaistausta	-0.038	0.009	-4.49	0
Korkein tutkinto toinen aste	0.013	0.006	2.1	0.036
Korkein tutkinto erikoisammattikoulu	0.025	0.021	1.2	0.231
Korkein tutkinto alin korkea-aste	0.080	0.008	9.87	0
Korkein tutkinto alempi korkeakoulu	0.115	0.008	15.11	0
Korkein tutkinto ylempi korkeakoulu	0.195	0.008	24.71	0
Korkein tutkinto tutkijakoulutus	0.352	0.013	28.1	0
Asuinpaikkakohtaiset muuttujat				
Asuminen pääkaupunkiseudulla	-0.245	0.005	-52.87	0
Rantarata asuinkunnassa	0.094	0.007	13.76	0
Päärata asuinkunnassa	-0.012	0.007	-1.68	0.092
VT1 asuinkunnassa	0.053	0.007	7.82	0
_cons	-1.981	0.015	-131.15	0
N	7027664			
LR chi2(25)	23797			
Prob>chi2	0.000			
Pseudo R2	0.047			

6.5 Pendelöinnin lopettamista selittävän ordered probit- mallin tuloksia

Pendelöinnin lopettamisesta estimoititiin vastaavanlainen ordered probit- malli kuin pendelöinnin aloittajistakin. Taulukossa 15 näkyy eri kynnsarvot ordered probit- mallille. Aineistossa on samanlaiset piirteet nähtävissä kuin pendelöinnin aloittaneidenkin aineistossa: pienet työmatkan muutokset ovat suuresti edustettuina ja aineiston häntä ei ole tasaisesti laskeva. Aineistoa on jokaisen kynnsarvon ryhmään tarpeeksi mallin ajamiseksi.

Taulukko 15. Aineiston jakautuminen eri kynnysväleille

Kynnysarvot	Frekvenssi
[0,-15]	1515976
[-15,-30)	59970
[-30,-60)	37490
[-60,-100)	12541
[-100,-300)	28962
Yhteensä	1654939

Taulukossa 16 on ordered probit- mallin tuloksia. Tilastollista merkitsevyyttä saa hyvin moni muuttaja. Vain toisen asteen tutkinnon suorittaneet eivät ole tilastollisesti merkitsevä ryhmä.

Eri kuntaan muuttajat ovat korkealla kertoimellaan tässäkin mallissa, mikä vahvistaa käsitystä, että muuttaminen on hyvin todennäköisesti yhteydessä jonkinlaiseen pendelöinnin muutokseen. Jokainen sosioekonominen asema, eli toisin sanoen työssä käyminen, vähentää pendelöinnin lopettamistodennäköisyyttä. Tämä on sinällään luonnollinen ajatus.

Asuminen pääkaupunkiseudulla saa negatiivisen kertoimen. Negatiivinen kerroin tuntuu olevan ominainen tälle ryhmälle jokaisessa mallissa. Pääkaupunkiseudulla asuminen tuntuisi kokonaisvaltaisesti vähentävän todennäköisyyttä pendelöinnin muutoksiin.

Tämä malli omalta osaltaan vahvistaa edellisten mallien antamaa tilannekuvaa. Pendelöinnin kannalta kiinnostavimmat ryhmät näiden tehtyjen probit- ja ordered probit- tarkastelujen valossa ovat suuren liikkuvuuden myötä korkeasti koulutetut, sekä eri kuntaan muuttajat ja miehet. Erityisen matalan liikkuvuuden osalta kiinnostavia ovat pääkaupunkiseudulla asuvat.

Taulukko 16. Ordered probit- mallin tuloksia pendelöinnin vähentäjistä

	Coef.	Std. Err.	z	P> z
Juokseva vuosi	0.012	0.000	25.87	0
Henkilökohtaiset muuttujat				
Ikä	-0.010	0.000	-65.35	0
Mies	0.132	0.003	42.37	0
Työttömyyskuukausia vuoden aikana	0.106	0.002	58.97	0
Aviossa	-0.013	0.003	-3.82	0
Omistusasuja	-0.023	0.007	-3.3	0.001
Vuokra-asuja	0.049	0.007	6.65	0
Käytössä olevat tulot tuhansina	-0.002	0.000	-13.79	0
Alle 7- vuotiaiden lasten lukumäärä	-0.012	0.002	-5.55	0
Tulojen muutos t-1 ja t välillä	-0.001	0.001	-1.96	0.049
Eri kuntaan muuttaja	0.796	0.004	179.95	0
Ylempi toimihenkilö	-0.340	0.006	-61.74	0
Alempi toimihenkilö	-0.383	0.005	-75.59	0
Työntekijä	-0.409	0.005	-75.21	0
Ulkomaalaistausta	-0.045	0.007	-6.63	0
Korkein tutkinto toinen aste	-0.005	0.005	-1.04	0.301
Korkein tutkinto erikoisammattikoulu	0.053	0.015	3.54	0
Korkein tutkinto alin korkea-aste	0.052	0.006	8.32	0
Korkein tutkinto alempi korkeakoulu	0.072	0.006	12.03	0
Korkein tutkinto ylempi korkeakoulu	0.144	0.006	22.62	0
Korkein tutkinto tutkijakoulutus	0.277	0.012	23.56	0
Asuinpaikkakohtaiset muuttujat				
Asuminen pääkaupunkiseudulla	-0.431	0.004	-110.55	0
Rantarata asuinkunnassa	-0.011	0.005	-1.98	0.048
Päärata asuinkunnassa	-0.077	0.005	-14.39	0
VT1 asuinkunnassa	-0.059	0.005	-10.93	0
/cut1	0.597	0.012		
/cut2	0.909	0.012		
/cut3	1.215	0.012		
/cut4	1.368	0.012		
N	1654939			
LR chi2(25)	77715			
Prob>chi2	0.0000			
Pseudo R2	0.060			

7 Pendelöinnin kestoa selittävien mallien tuloksia

7.1 Coxin mallin tuloksia

Coxin malli estimoitiin 60km selviytymisaineistolla sekä 100km selviytymisaineistolla. Selviytymisaineiston tulkintana on se, että jokainen kaukopendelöijä asettaa itsensä alttiiksi riskille lopettaa kaukopendelöinti. Ne, jotka eivät kaukopendelöi, eivät ole riskinalaisia havaintoja. Henkilö voi aloittaa ja lopettaa kaukopendelöinnin moneen kertaan tarkastelujaksolla. Coxin mallin regressiokertoimet kertovat siis muuttujan kaukopendelöinnin lopettamiselta suojaavalta vaikutukselta tai kaukopendelöinnin lopettamiselle altistavalta vaikutukselta. Kiinnostuksen kohteena erityisesti on edellisten lukujen probit- malleissa esille tulleet ryhmät. Muuttujien vaaran asteesta voidaan tulkita pendelöintiä ylläpitäviä tekijöitä, sekä pendelöinnin lopettavia tekijöitä. Tässä tulkinnassa yli 1 olevat vaaran asteet ovat kaukopendelöintiä lopettavia tekijöitä, sekä alle 1 olevat vaaran asteet kaukopendelöintiä ylläpitäviä tekijöitä.

Aikaisemmassa probit- tarkastelussa pystyttiin käyttämään koko aineistoa mallin ajamisessa, mutta selviytymismallinnuksessa havaintoja ovat vain tietyllä periodilla kaukopendelöivät. Siten myös aineisto vähenee radikaalisti. Kuitenkin aineistoa on verraten runsaasti, jotta jotain päätelmiä on mahdollista tehdä.

Taulukossa 17 on kuvattu Cox- mallin vaaran asteet 60 kilometrin kynnyksarvolla, sekä taulukossa 18 on kuvattu vastaavat 100 kilometrin kynnyksarvolla. Mallit ovat tilastollisesti merkittäviä. Muuttujat saavat yleisesti heikosti tilastollista merkitsevyyttä tässä tarkastelussa. Osasyynä on edellä mainittu selviytymisaineiston pienuus. Kuitenkin silti kiinnostaviakin uusia havaintoja löytyy.

Ainoat tilastollista merkitsevyyttä saavat muuttujat ovat ikä, mies ja käytössä olevat tulot. Tilastollisen merkitsevyyden rajalla ovat työttömyyskuukaudet, vuokra-asuminen, ja tulojen muutos. Ikä toimii pendelöinnin lopettamiselle altistavana tekijänä. Tämä voi olla vastoin aikaisempia tuloksia Alasalmi ym. (2020), jossa ikä nähdään pendelöintiä selittävänä tekijänä. Työttömyyskuukaudet näyttäisivät olevan kaukopendelöintiä ylläpitävä tekijä ja vuokra-asuminen kaukopendelöintiä lopettava tekijä. Käytössä olevat tulot eivät vaikuta kaukopendelöinnin kestoon juurikaan mitään. Kiinnostavaa kuitenkin on havaita, että tulojen muutos kuitenkin on kaukopendelöintiä ylläpitävä tekijä. Eli hyvä palkkakehitys voisi ylläpitää kaukopendelöinnin jatkamista.

Taulukko 17. Cox- mallin vaaran asteen estimaatit 60 kilometrin kynnsarvolla

	Haz. Ratio	Std. Err.	z	P> z
Henkilökohtaiset muuttujat				
Ikä	1.010	0.001	7.320	0.000
Mies	0.937	0.025	-2.420	0.016
Työttömyyskuukausia vuoden aikana	0.970	0.015	-1.960	0.050
Aviossa	0.971	0.027	-1.050	0.292
Omistusasuja	1.052	0.065	0.820	0.410
Vuokra-asuja	1.045	0.069	0.670	0.503
Käytössä olevat tulot tuhansina	1.003	0.001	4.170	0.000
Alle 7- vuotiaiden lasten lukumäärä	0.969	0.018	-1.670	0.094
Tulojen muutos t-1 ja t välillä	0.948	0.026	-1.930	0.054
Eri kuntaan muuttaja	0.950	0.035	-1.390	0.166
Ylempi toimihenkilö	0.988	0.044	-0.270	0.785
Alempi toimihenkilö	1.054	0.048	1.170	0.243
Työntekijä	1.053	0.053	1.030	0.302
Ulkomaalaistausta	1.039	0.075	0.540	0.591
Korkein tutkinto toinen aste	0.999	0.045	-0.010	0.988
Korkein tutkinto erikoisammattikoulu	0.877	0.108	-1.060	0.289
Korkein tutkinto alin korkea-aste	0.916	0.049	-1.630	0.102
Korkein tutkinto alempi korkeakoulu	0.988	0.051	-0.240	0.809
Korkein tutkinto ylempi korkeakoulu	0.949	0.050	-1.010	0.315
Korkein tutkinto tutkijakoulutus	0.936	0.075	-0.830	0.405
Asuinpaikkakohtaiset muuttujat				
Asuminen pääkaupunkiseudulla	0.965	0.030	-1.140	0.255
Rantarata asuinkunnassa	1.030	0.038	0.810	0.418
Päärata asuinkunnassa	1.062	0.058	1.100	0.269
VT1	1.012	0.038	0.310	0.757
Kaukopenelöijä	6869			
Epäonnistumisia (engl. <i>failures</i>)	7210			
Aika riskin alaisuudessa	19483			
N	7526			
Wald chi2(24)	153.3753			
Prob > chi2	0.001			

100 kilometrin kynnsarvolla tilastollisesti merkitsevät muuttujat vähenevät entisestään. Nyt tilastollisesti merkitseviä muuttujia on vain ikä, mies- sukupuoli, käytössä olevat tulot ja tulojen muutos. Juurikaan mitään muutosta ei tapahtunut verrattuna 60 kilometrin kynnsarvolla olleeseen malliin. Molempien mallien tapauksessa yleinen havainto on erittäin lähellä 1 olevat vaaran asteet. Tämän mukaan vaaran aste ei ole kovinkaan riippuvainen mistään mallissa käytetystä muuttujasta.

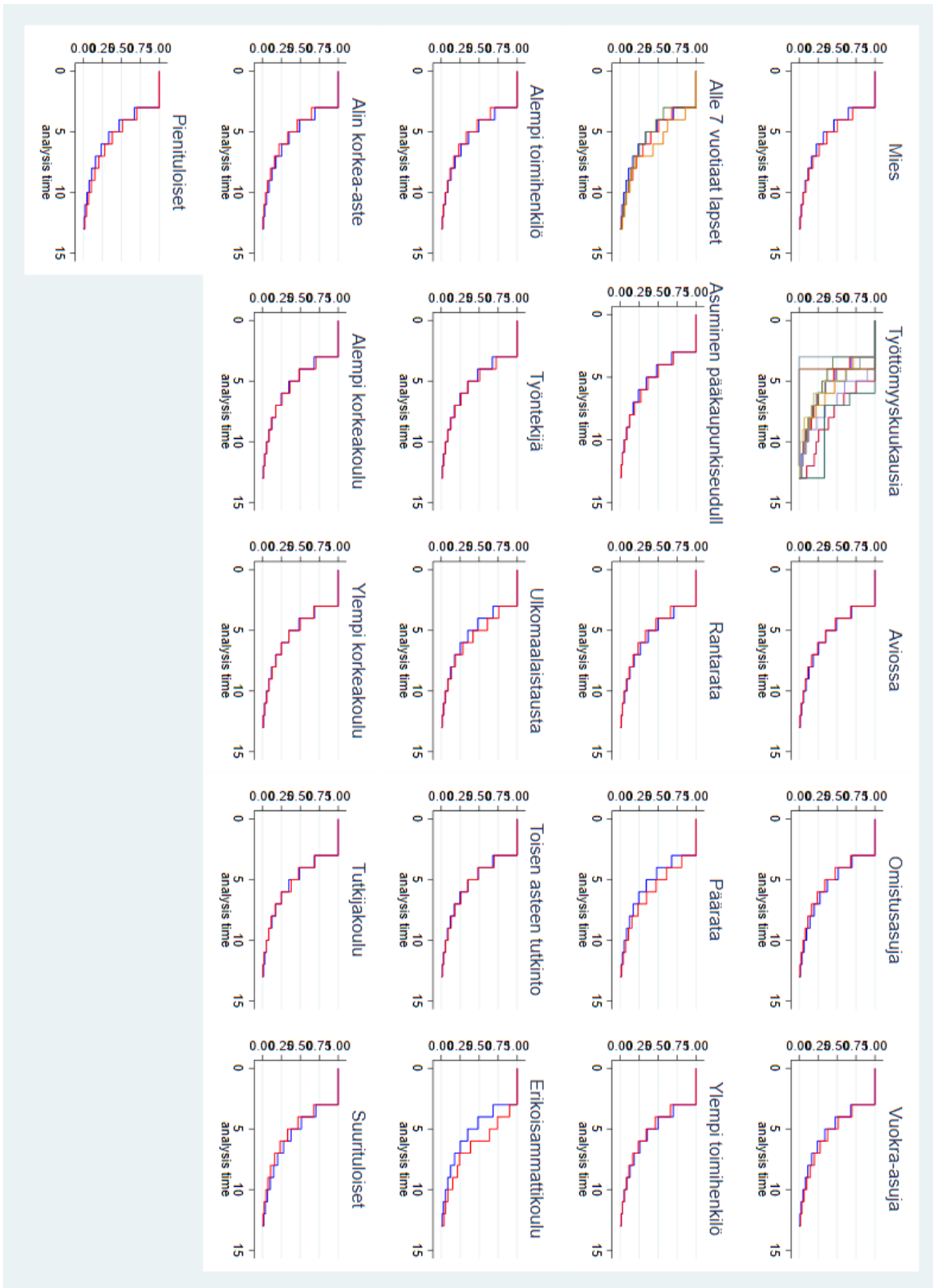
Taulukko 18. Cox mallin vaaran asteen estimaatit 100 kilometrin kynnyksarvolla

	Haz. Ratio	Std. Err.	z	P> z
Henkilökohtaiset muuttujat				
Ikä	1.011	0.002	6.220	0.000
Mies	0.919	0.032	-2.450	0.014
Työttömyyskuukausia vuoden aikana	0.981	0.019	-0.980	0.329
Aviossa	0.943	0.033	-1.670	0.095
Omistusasuja	1.008	0.077	0.110	0.916
Vuokra-asuja	0.983	0.082	-0.200	0.839
Käytössä olevat tulot tuhansina	1.003	0.001	3.530	0.000
Alle 7- vuotiaiden lasten lukumäärä	0.983	0.024	-0.690	0.491
Tulojen muutos t-1 ja t välillä	0.926	0.031	-2.290	0.022
Eri kuntaan muuttaja	0.941	0.047	-1.210	0.225
Ylempi toimihenkilö	0.912	0.052	-1.620	0.105
Alempi toimihenkilö	0.988	0.057	-0.200	0.840
Työntekijä	0.987	0.064	-0.200	0.842
Ulkomaalaistausta	1.013	0.093	0.140	0.887
Korkein tutkinto toinen aste	0.960	0.056	-0.690	0.491
Korkein tutkinto erikoisammattikoulu	0.937	0.157	-0.390	0.695
Korkein tutkinto alin korkea-aste	0.889	0.062	-1.690	0.091
Korkein tutkinto alempi korkeakoulu	0.981	0.065	-0.290	0.773
Korkein tutkinto ylempi korkeakoulu	0.936	0.063	-0.980	0.326
Korkein tutkinto tutkijakoulutus	0.911	0.087	-0.980	0.327
Asuinpaikkakohtaiset muuttujat				
Asuminen pääkaupunkiseudulla	0.971	0.037	-0.770	0.443
Rantarata asuinkunnassa	1.014	0.054	0.270	0.791
Päärata asuinkunnassa	0.984	0.069	-0.230	0.818
VT1	1.006	0.054	0.110	0.909
Kaukopenseläisiä	4251			
Epäonnistumisia (engl. <i>failures</i>)	4476			
Aika riskin alaisuudessa	11975			
N	4682			
Wald chi2(24)	110.5551			
Prob > chi2	0.001774			

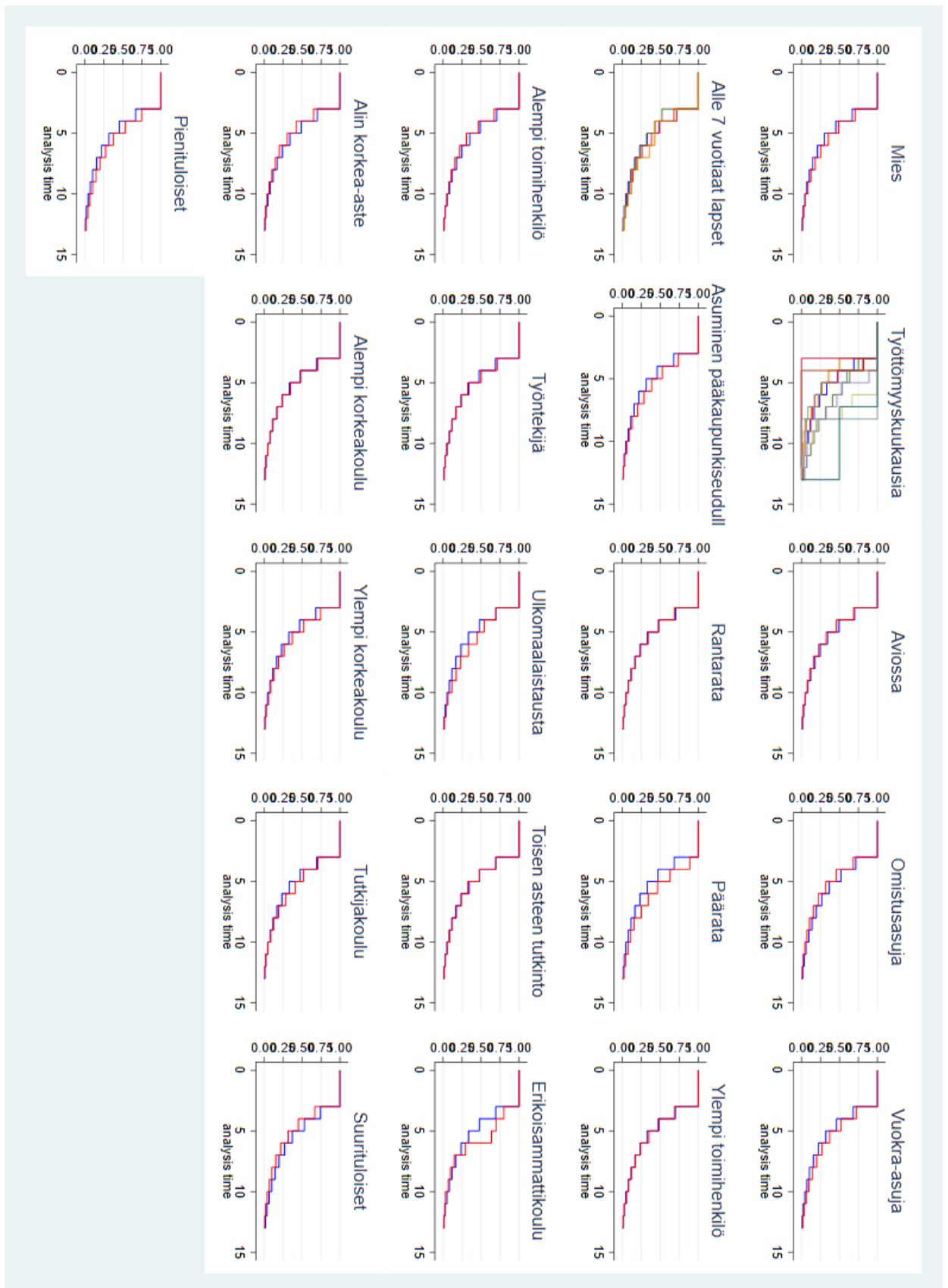
7.2 Coxin mallin oletusten tarkastelua

Tässä tutkielmassa käytetyn aineiston Kaplan-Meier- kuvaajat on esitetty Kuviossa 4 ja 5. Verrokkiryhmä on kuvattu sinisellä käyrällä. Muuttujista on valittu kuvaajiin ne, joilla Kaplan-Meier- käyrien visuaalinen tarkastelu on mielekästä. Esimerkiksi iän tapauksessa tarkastelua olisi voitu tehdä, mutta kategorioita on niin useita, ettei visuaalisesta tarkastelusta selvinne juuri mitään. Laskenta-ajan säästämiseksi on tulo-

muuttuja jaoteltu mediaanitulojen mukaan kahteen kategoriaan: suurituloiset, sekä pienituloiset. Coxin mallin suhteellisten riskien oletusten vastaisia muuttujia näyttäisi olevan reilusti työttömyyskuukaudet. Kaplan-Meier- käyrät risteävät ja ylittävät toisiaan. Alle 7 vuotiaat lapset ovat kyseenalaisia molemmissa kuvioissa. Muissa muuttujissa ei ole juurikaan suurta risteävyyttä. Jonkinlainen erkaneminen tapahtuu tulomuuttujissa, sekä erikoisammattikoulu- muuttujassa.



Kuvio 4. Kaplan-Meier käyrät 60 kilometrin selviytymisaineistossa muuttujittain.



Kuvio 5. Kaplan-Meier käyrät 100 kilometrin selviytymisaineistossa muuttujittain

Kaplan-Meier käyrien eriävyyttä voidaan tarkastella sekä visuaalisesti, että tilastollisella logrank- testillä. Logrank testi on mahdollista tehdä myös niille muuttujille, joiden visuaalinen tarkastelu ei olisi kovinkaan mielekäästä.

Taulukossa 19 esitetyistä logrank- testin tuloksista voidaan tulkita, että ne noudattelevat Kaplan-Meier kuvaajista visuaalisesti saatavaa tietoa: korkean p-arvon saaneiden muuttujien Kaplan-Meier- käyrät ovat yhteneväiset, ja matalan p-arvon saaneiden erkanevat toisistaan. Logrank- testi antaa visuaaliselle tarkastelulle tilastollisen vahvistuksen. Tämä antaa tietoa siitä, että mitkä muuttujat voisivat olla kiinnostavia jatkotutkimuksen kannalta selviytymisanalyysin perusteella, sekä antaa osviittaa myös mahdollisista suhteellisten riskien oletusten rikkoontumisesta. Tässä tarkasteluissa hyvin karkealla tulojen jaottelulla on saatu ero selviytymiskäyrien välille, mikä tarkoittaa, että suurituloiset näyttäisivät ylläpitävän pendelöintiä vähemmän verrokkiryhmään verrattuna, kun taas pienituloiset enemmän.

Taulukko 19. Logrank- testin tulokset

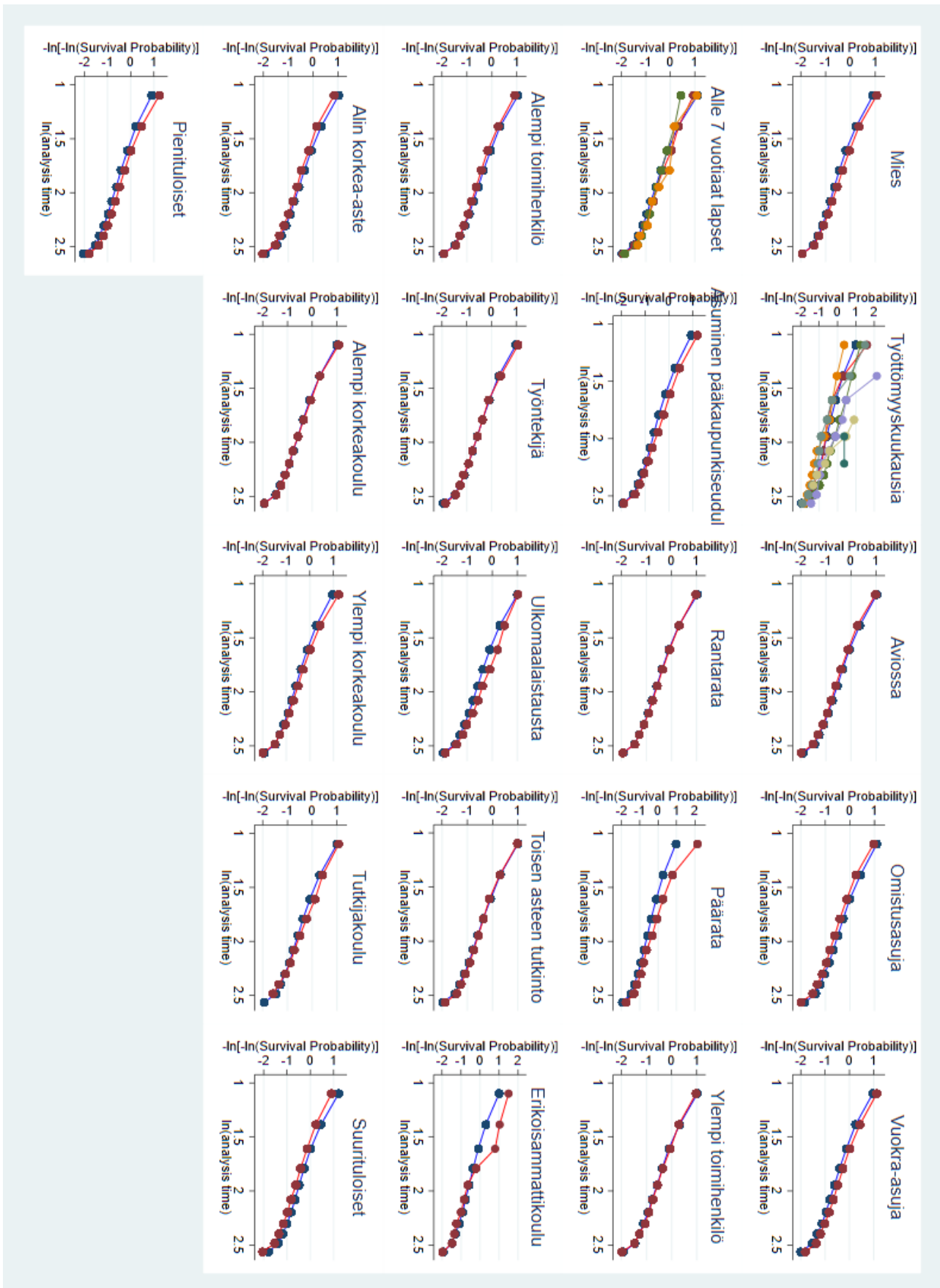
	p-arvo 60km	p-arvo 100km
Henkilökohtaiset muuttujat		
Ikä	0	0
Mies	0.462	0.429
Työttömyyskuukaudet	0.060	0.011
Aviossa	0.002	0.068
Omistusasuja	0	0
Vuokra-asuja	0	0
Alle 7- vuotiaiden lasten lukumäärä	0	0.003
Ylempi toimihenkilö	0.355	0.799
Alempi toimihenkilö	0.650	0.708
Työntekijä	0.192	0.678
Ulkomaalaistausta	0.405	0.337
Korkein tutkinto toinen aste	0.012	0.006
Korkein tutkinto erikoisammattikoulu	0.086	0.838
Korkein tutkinto alin korkea-aste	0.008	0.122
Korkein tutkinto alempi korkeakoulu	0.255	0.572
Korkein tutkinto ylempi korkeakoulu	0.413	0.665
Korkein tutkinto tutkijakoulutus	0.064	0.174
Pienituloiset	0	0
Suurituloiset	0	0
Asuinpaikkakohtaiset muuttujat		
Asuminen pääkaupunkiseudulla	0.416	0.073
Rantarata asuinkunnassa	0.810	0.612
Päärata asuinkunnassa	0.229	0.017
VT1	0.788	0.640

Log-log- käyrät ovat Kaplan-Meier- käyrien kanssa eräs visuaalinen työkalu suhteellisten vaarojen oletusten tarkastelemiseksi. Jotta suhteellisten vaarojen oletus pitäisi paikkaansa, log-log käyrien tulisi kulkea samansuuntaisesti toisiinsa nähden

Kuviossa 6 ja 7 on kuvattuna 60 kilometrin, sekä 100 kilometrin selviytymisaineiston log-log- käyrät. Log-log- käyrät vahvistavat Kaplan-Meier- käyristä saatua tietoa. Kaplan-Meier käyrien perusteella ongelmallisia muuttujia olivat erkanemista tarkastellen tulomuuttujat ja erikoisammattikoulu- muuttuja. Risteävyyttä ajatellen työttömyyskuukaudet, sekä alle 7- vuotiaat lapset.

Työttömyyskuukaudet ovat vaikeat tulkittavat log-log- käyristä usean käyrän vuoksi. Kuitenkin on havaittavissa, että käyrät kulkevat jokseenkin rinnakkain, mutta risteämistä tapahtuu analyysin alkuvaiheessa, jolloin suurin osa aineistosta on vielä selviytynyttä. Alle 7- vuotiaat lapset- muuttujan log-log- käyrä näyttää pientä erkanemista alussa kohtisuoruudesta, sekä alussa tapahtuu risteäminen, mutta tämä on

jälleen hyvin tulkinnanvarainen rajatapaus. Erikoisammattikoulu- kuviassa näkyy myös selviytymisanalyysin alkuvaiheessa tapahtuva erkaneminen. Kuitenkin erikoisammattikoulun käyneitä on aineistossa oletetusti hyvin vähän, jolloin analyysin kannalta huolta ei kannata asiasta kantaa. Tulomuuttajat, joiden käyrät Kaplan-Meier-tarkastelussa erkanivat toisistaan, pysyvät log-log- tarkastelussa hyvin saman suuntaisina. Siten oletettavasti mallin oletukset pitävät myös niiden muuttujien tapauksessa.



Kuvio 7. 100 kilometrin selviytymisaineiston log-log käyrät muuttujittain

Suhteellisten riskien oletusta on hyvä tarkastella myös ajassa olevien jaksottaisten trendien ja syklien osalta. Tähän tarkasteluun voidaan käyttää Schoenfeld- testiä. Schoenfeld- testissä testataan koko malli aikatrendien varalta, sekä yksittäiset muuttujat. Taulukossa 20 on kuvattuna Schoenfeld- testin p-arvot. 100 kilometrin kynnysarvon malli, kuten mallin muuttujatkaan eivät saa tilastollista merkitsevyyttä. Tämä tarkoittaa, että nollahypoteesi hyväksytään kaikille muuttujille. 60 kilometrin malli on tässä tarkastelussa ongelmallinen. Siten siis sellaista aikatrendiä tai syklittäisyyttä ei ole mikä voisi olla Coxin mallin oletusten vastainen 100 kilometrin mallissa, mutta 60 kilometrin mallissa sellainen voi olla.

Taulukko 20. Schoenfeld- testin tulokset 60kilometrin sekä 100 km mallille muuttujittain

	p- arvo 60km	p- arvo 100km
Malli	0.026	0.090
Henkilökohtaiset muuttujat		
Ikä	0.008	0.045
Mies	0.036	0.022
Työttömyyskuukausia vuoden aikana	0.968	0.992
Aviossa	0.415	0.905
Omistusasuja	0.399	0.496
Vuokra-asuja	0.440	0.319
Käytössä olevat tulot tuhansina	0.024	0.018
Alle 7- vuotiaiden lasten lukumäärä	0.812	0.659
Tulojen muutos t-1 ja t välillä	0.648	0.214
Eri kuntaan muuttaja	0.140	0.236
Ylempi toimihenkilö	0.047	0.343
Alempi toimihenkilö	0.020	0.043
Työntekijä	0.034	0.402
Ulkomaalaistausta	0.583	0.745
Korkein tutkinto toinen aste	0.693	0.760
Korkein tutkinto erikoisammattikoulu	0.895	0.920
Korkein tutkinto alin korkea-aste	0.504	0.557
Korkein tutkinto alempi korkeakoulu	0.982	0.752
Korkein tutkinto ylempi korkeakoulu	0.262	0.158
Korkein tutkinto tutkijakoulutus	0.082	0.058
Asuinpaikkakohtaiset muuttujat		
Asuminen pääkaupunkiseudulla	0.352	0.330
Rantarata asuinkunnassa	0.184	0.182
Päärata asuinkunnassa	0.639	0.995
VT1	0.834	0.267

8 Yhteenveto

Tässä tutkielmassa pyrittiin etsimään ratkaisuja seuraaviin tutkimuskysymyksiin:

1. Mitkä taustatekijät selittävät kaukopendelöintiä?
2. Mitkä taustatekijät voivat laukaista tai lopettaa kaukopendelöinnin?
3. Mitkä taustatekijät toimivat kaukopendelöintiä ylläpitävänä tekijänä?

Ensimmäiseen tutkielmalle asetettuun tutkimuskysymykseen probit- malli antoi erittäin hedelmällistä tietoa rekisteriaineistosta. Probit- sekä ordered- probit mallit toimivat erittäin hyvin kaukopendelöijien profiloinnissa ja ne antoivat arvokasta tietoa.

Kenties kaikkien mallinnusten perusteella voisi todeta kaukopendelöijän profiilin olevan mies, joka on koulutustasoltaan vähintään korkeakoulutunut, mutta erityisesti tutkijakoulutuksen saanut. Lisäksi hän on muuttanut juuri paikkakunnalle ja ei asu pääkaupunkiseudulla. Koulutuksen yhteys suurempaan pendelöintialttiuteen on yhtenevää Helminen & Ristimäki (2007) löydösten kanssa, mikäli oletamme korkean koulutuksen mahdollistavan yleisesti etätöiden tekemisen kotona. Yleisesti pendelöintiä selitti ordered probit- mallin tarkastelussa luonnollisesti työssä käyminen. Tässäkin tarkastelussa pääkaupunkiseudun asukkaat edustivat erittäin alhaista liikkuvuutta. Eli pääkaupunkiseudun asukkaat kaukopendelöinnin lisäksi yleisemminkin pendelöivät vähän. Tämä vahvistaa Alasalmi ym. (2020) analyysiä pääkaupunkiseudun asukkaiden hyvin matalasta liikkuvuudesta.

Liikenneyhteyksistä kiinnostavana tarkastelussa näyttäytyi Rantarata. Rantaradan henkilöpaikan omaavissa kunnissa asuminen lisäsi todennäköisemmin 60 kilometrin kynnyksarvolla pendelöintiä, kuin 100 kilometrin. Kuitenkin VT1 regressiokerroin oli 100 kilometrin kynnyksarvolla reilusti positiivinen, kun taas 60 kilometrin kynnyksarvolla reilusti negatiivinen. Näyttäisi siis, että kunnissa, joiden läpi VT1 kulkee, asuu pidemmällä pendelöiviä ihmisiä kuin kunnissa, joissa on Rantaradan henkilöliikennepaikka. Tämä on hyvin epäintuitiivinen löydös, sillä olisi oletettavaa, että lyhyempiä matkoja kuljettaisiin autolla valtateitä pisin ja pidempiä matkoja junilla. Pääkaupunkiseudun hyvin vähäinen pendelöinti saattaa selittää osaltaan näitä havaintoja.

Toiseen tutkielmalle asetettuun tutkimuskysymykseen löytyi myös selittäviä tekijöitä. 60 kilometrin kynnsarvon sekä 100 kilometrin kynnsarvon mallien päälinja oli kutakuinkin samanlainen. Tässä tarkastelussa eri kuntaan muuttajat olivat hyvin mielenkiintoinen ryhmä. Muuttaminen hyvin suurella todennäköisyydellä selittää sekä kaukopenelöinnin aloittamista ja lopettamista. Muuttaminen on kiinnostava muuttaja siinä mielessä, että se ei ole tutkijakoulutuksen kaltainen pysyvä tila. Muuttaja on väliaikainen tila, ellei muuta vuosittain. Muuttajat korostuivat myös ordered probit-tarkastelussa.

Pääkaupunkiseudulla asuminen vaikutti negatiivisesti sekä kaukopenelöinnin, että ordered probit- mallissa kaikenlaisen pendelöinnin aloittamiseen, että lopettamiseen. Tämä vahvistaa entisestään kuvaa pääkaupunkiseudun asukkaiden matalasta liikkuvuudesta.

Kolmanteen tutkimuskysymykseen pyrittiin hakemaan vastausta survival- malleilla. Vaikka mallit olivat tilastollisesti merkitseviä, muuttujista oli vaikea saada merkitsevyyttä ulos. Ongelmia oli myös Coxin mallin taustaoletusten rikkoontumisten kanssa eräiden muuttajien tapauksessa. Malli tuotti myös hyvin lähellä 1 olevia vaaran asteita, minkä voi tulkita, että tässä tutkielmassa käytetyillä muuttujilla ei ole juuri millaistaakaan yhteyttä kaukopenelöinnin lopettamisalttiuteen. Vaikka tässä tutkielmassa survival- mallinnus osoittautui vienoksi pettymykseksi, yleisesti lääketieteessä sekä biometriikassa käytetyllä survival- analyysillä kuitenkin voisi olla mielenkiintoisia sovellutuskohteita myös yhteiskuntatieteissä.

Tulokset antavat lisää tietoa millaisiin ihmisiin kohdistuu erilaisten liikennehankkeiden hyödyt. Erityisesti kaksi kiinnostavaa ryhmää nousee esiin: tutkijakoulutetut sekä eri kuntaan muuttajat. Tutkijakoulutuksen käyneet toimivat usein erilaisissa asiantuntijatehtävissä. Hyvät pendelöintimahdollisuudet voisivat mahdollisesti toimia vetävänä tekijänä erilaisille asiantuntijoille pysyä tai muuttaa Suomeen. Lisäksi yleisesti tutkijakoulutettujen suuri liikkuvuus antaa positiivista tietoa siitä, että pitkän koulutuksen läpikäyneet pystyvät ottamaan työpaikkoja vastaan pitkienkin matkojen päästä. Tämä on oleellinen havainto yhteiskunnan koulutusresurssien käytön kannalta: jos tutkijakoulutetut eivät liikkuisi, se voisi kuvastaa myös laajempaa ongelmaa koulutuksen hukkaan menemisestä ja työllistymisestä matalamman osaamistason työhön, kuin mihin koulutus mahdollistaisi.

Suuri aineisto yli 7 miljoonalla datapisteellä tarjoaisi mahdollisuudet paljon spesifimpään probit- analyysiin, mitä tämän tutkielman resursseilla oli mahdollista tehdä. Mahdollinen jatkotutkimuksen aihe voisi olla tässä tutkielmassa tehdyn analyysin tarkentaminen. Esimerkiksi ilmeisiä puutteita analyysiin on ihmisten toimialojen puuttuminen. Jotkut toimialat ovat sen luonteisia, että niissä yrityksen rekisteriaineistosta löytyvä kotipaikka ei välttämättä kerro kovin hyvin missä tosiasiallinen työskentely tapahtuu. Tämän lisäksi myös toimialojen lisääminen yleisesti tarkentaisi analyysiä.

Muuttajat olivat tämän tutkielman valossa jatkotutkimuksen kannalta mielestäni kiinnostava ryhmä. Muuttajat korostuivat sekä pendelöijinä, pendelöinnin aloittajina, että lopettajina kuitenkin omaten keskiarvoa vähemmän pendelöintiperiodeja. Muuttamisen ja pendelöinnin vaikutusmekanismeista olisi mielenkiintoista saada uutta tietoa. Kiinnostavia kysymyksiä on, liittyykö muuttajien korostuminen työpaikan vaihtoon ja miten nämä ajallisesti järjestyvät. Muutetaanko siis ensin toisaalle ja pendelöidään entiseen työpaikkaan kauas siihen saakka, kunnes löydetään työ uuden kotipaikan läheltä, vai pendelöidäänkö kauas työpaikkaan ja muutetaan työpaikan lähelle. Tällä voisi olla merkitystä kuntien houkuttelevuuden kannalta: kannattaako kunnan panostaa hyviin työmahdollisuuksiin vai yleiseen viihtyvyyteen ja sosiaalisten ongelmien vähentämiseen. Mikäli muuttamispäätös syntyy työmahdollisuuksista, voisi sosiaaliset ongelmat vähentyä itsestään, kun alueelle muuttaa työssä käyvää väkeä. Jos muuttamispäätös syntyy kunnan yleisen maineen perusteella, silloin oletettavasti myös työmahdollisuuksia kuntaan syntyy ihmisten muuttaessa alueelle.

Lähteet

- Alasalmi J, Busk H., Holappa, V., Huovari, J., Härmälä, V., Kotavaara, O., Lehtonen, O., Muilu, T., Rusanen, J. ja Vihinen, H. (2020) *Työn ja työvoiman alueellinen liikkuvuus ja monipaikkainen väestö* Valtioneuvoston kanslia, Helsinki.
- Aalen, O., Borgan, Ö. & Gjessing, H. ((2008) *Survival and event history analysis: Process point of view* Springer Science & Business Media, New York: USA
- Burd, C., Burrows, M. & McKenzie, B. (2021) Travel time to Work in the United States: 2019. *American Community Survey Reports*.
- Champion, T. et al. (2009) Migration and Longer-Distance Commuting in Rural England. *Regional Studies*. 43 (10), 1245–1259.
- Cramer, J. S. (1986) *Econometric Applications of Maximum Likelihood Methods*. Cambridge: Cambridge University Press.
- Cuff, D. (2011) Los Angeles: Urban Development in the Postsuburban Megacity. In: Sorensen, A., Okata, J. (eds) *Megacities. Library for Sustainable Urban Regeneration*, vol 10. Springer, Tokyo.
- The Ford European Commuter Survey (2015) *For Europeans, the Journey to Work Causes More Stress Than Their Actual Jobs (or even the Dentist), New Ford Survey Shows*.
<https://media.ford.com/content/fordmedia/feu/en/news/2015/04/27/for-europeans--the-journey-to-work-causes-more-stress-than-their.html> Haettu: 1.5.2024
- Hess, K. (1995) Graphical methods for assessing violations of the proportional hazards assumption in Cox regression. *Statistics in Medicine*. 14, 1707-1723.
- Holm, P., Nivalainen, S. & Volk, R. (2008) Työvoiman alueellisen liikkuvuuden kannattavuus – OSA 1. *Valtioneuvoston kanslian julkaisusarja 1/2008*. Helsinki, Yliopistopaino.
- Helminen, V. & Ristimäki, M. (2007) Relationships between commuting distance, frequency and telework in Finland. *Journal of Transport Geography*. 15 (5), 331–342.
- Kahneman, D. & Krueger, A. B. (2006) Developments in the Measurement of Subjective Well-Being. *The Journal of Economic Perspectives*. 20 (1), 3–24.

- Limtanakool, N (2006) On the participation in medium- and long-distance travel: A decomposition analysis for the UK and the Netherlands. *Tijdschrift Voor Economische en Sociale Geografie*. 97 (4), 389–404.
- Maoh, H. & Tang, Z. (2012) Determinants of normal and extreme commute distance in a sprawled midsize Canadian city: evidence from Windsor, Canada. *Journal of Transport Geography*. 25, 50–57.
- McCullagh, P. & Nelder, J. A. (1983) *Generalized Linear Models*. 2nd edition. 37. Boca Raton: CRC Press.
- Marion, B. & Horner, M. W. (2007) Comparison of Socioeconomic and Demographic Profiles of Extreme Commuters in Several U.S. Metropolitan Statistical Areas. *Transportation Research Record 2013*. (1), 38–45.
- Monte, F., Redding, S. & Rossi-Hansberg, E. (2018) Commuting, Migration, and Local Employment Elasticities. *American Economic Review 2018*. 108(2), 3855-3890.
- Owens III, R., Rossi-Hansberg, E. & Sarte, P. (2020) Rethinking Detroit. *American Economic Journal: Economic Policy 2020*. 12 (2), 258-305.
- Paumgarten, N. (2007) There and back again. *The New Yorker*.
<https://www.newyorker.com/magazine/2007/04/16/there-and-back-again>
 Haettu 15.4.2024
- Rickard, J. M. (1988). Factors Influencing Long-Distance Rail Passenger Trip Rates in Great Britain. *Journal of Transport Economics and Policy*, 22(2), 209–233.
- Schoenfeld, D. (1982) Partial residuals for the proportional hazards regression model. *Biometrika*, 69(1), 239-241.
- Stock, J & Watson, M (2007) *Introduction to Econometrics*, Pearson education, Boston
- Keva, J. (2020) Työmatkat ovat pidentyneet mutta puolet on hyvin pyöräiltävissä.
<https://www.stat.fi/tietotrendit/artikkelit/2020/tyomatkat-ovat-pidentyneet-mutta-puolet-on-hyvin-pyorailtavissa/?listing=simple> Haettu 14.4.2024
- Vani Kant, B. (2002) *Logit and Probit: Ordered and Multinomial Models*. 138. Los Angeles: SAGE Publications Inc.
- Vincent-Geslin, S. & Ravalet, E. (2016) Determinants of extreme commuting: Evidence from Brussels, Geneva and Lyon. *Journal of Transport Geography*. 54, 240-247
- Yle (2015) Suomalaisen työmatka pitenee metrin joka työpäivä-Kantahämäläiset suhaavat kauimmaksi. <https://yle.fi/a/3-7877400> Haettu 14.4.2024

