

Matti Mäenpää

Logistiikkaketjun tilastollinen mallinnus
Suomen vaatetuonnissa

Tilastotieteen
pro gradu – tutkielma
15.04.2008

Jyväskylän yliopisto
Matematiikan ja tilastotieteen laitos
Jyväskylä

Tiivistelmä

Mäenpää, Matti Mikko Matias

Tilastotieteen pro gradu-tutkielma / Matti Mäenpää

Jyväskylän yliopisto, 2008

50 s.

Pro gradu-tutkielma:

Logistiikkaketjun tilastollinen mallinnus Suomen vaatuonnissa.

Tässä tutkielmassa tutkitaan Suomen sisäkaupan vaatuontia. Tämä pro gradu-tutkielma perustuu tullin tuontitilastoihin. Tutkielmassa perehdytään Suomen sisäkaupan vaatuontiin vuodelta 2005. Tutkielmassa perehdytään myös logistisen regressioanalyysin teoriaan ja sitä käytetään työkaluna mallinnettaessa tutkielman tuloksia.

Suomen sisäkaupan alkuperämaajakauman on epäilty olevan epäluotettava. Tämän tilastotieteen pro gradu -tutkielman tarkoituksena on tarkastella missä määrin on mahdollista käytettävissä olevien tietojen perusteella estimoida luotettavasti Suomen sisäkaupan alkuperämaajakaumaa vaatuonnin osalta vuodelta 2005. Tutkielmassa tarkoituksena on tuottaa esimerkinomaisesti korjattu alkuperämaajakauma vaatuonnista Suomeen Ruotsin kautta. Tämän kaltaista tutkimusta ei ole aikaisemmin tehty.

Keskeisenä tuloksena tutkielmassa voidaan pitää sitä, että käytettävissä olevilla tiedoilla alkuperäinen alkuperämaajakauma tulee muuttumaan melko vähän. Näin ollen ei ole tarpeeksi vahvaa näyttöä alkuperämaatiedon laadusta. Tutkielman tulokset siis antavat vain apua ongelmaa tutkittaessa.

Alkuperämaatiedon keräämismenetelmiä olisi hyvä kehittää, jos tietoa on tarkoitus tilastoida myös tulevaisuudessa. Tutkielman kannalta ongelmana voidaan pitää sitä, että alkuperämaatiedon ilmoittaminen ei ole pakollinen tulli-ilmoituksissa.

Uusi mallinnettu alkuperämaajakauma muuttaa lähtötilanteen jakaumaa 14,2 miljoonalla eurolla, kun koko uudelleen mallinnettavan osuuden arvo oli 79,5 miljoonaa euroa, mikä on 13,5 % koko sisäkaupan vaatetuonnista.

Avainsanat: Logistiikkaketju, sisäkauppa, tullitilasto, alkuperämaa, lähtömaa, logistinen regressio.

SISÄLLYSLUETTELO

1	JOHDANTO.....	8
1.1	TUTKIELMAN TAUSTAA	8
1.2	TIEDONANTOVELVOLLISET TULLITILASTOSSA	8
1.3	ULKOMAANKAUPPATILASTO.....	8
1.4	MÄÄRITYKSIÄ TUONNISTA JA VIENNISTÄ.....	9
1.5	TUTKIMUSONGELMASTA	10
1.6	MIKSI ALKUPERÄMAATIETO ON HYÖDYLLINEN	10
1.7	TUTKIELMAN ETENEMINEN JA MENETELMÄT	12
1.8	EUROOPAN UNIONIN SISÄKAUPAN PERUSASETUS.....	14
1.9	TUTKIELMAN TAVOITTEET TIIVISTETYSTI	15
2	KESKEISET KÄSITTEET	17
3	SUOMEN SISÄ- JA ULKOKAUPPA.....	19
3.1	SUOMEN SISÄKAUPPA-AINEISTO (INTRASTAT)	19
3.2	SUOMEN ULKOKAUPPA-AINEISTO (EXTRASTAT)	19
3.3	EUROSTATIN AINEISTO KOKO EUROOPAN UNIONIN ALUEELLE.....	19
4	AINEISTON RAJAUS.....	20
4.1	TUTKIMUSAINEISTO JA AINEISTON RAKENNE.....	20
4.2	VAAATERYHMIEN YHDISTÄMINEN	21
4.3	SELITTÄVÄ OSA AINEISTOSTA	22
4.4	SELITETTÄVÄ OSA AINEISTOSTA.....	25
5	TILASTOLLISET MENETELMÄT	27
5.1	LOGISTINEN REGRESSIO	27
5.2	SELITYSASTEEN MÄÄRITELMÄ LOGISTISESSA REGRESSIOSSA	28
5.3	LOGISTINEN REGRESSIO, RIIPPUMATTOMAT HAVAINNOT	29
5.4	LOGISTINEN REGRESSIO, RIIPPUVAT HAVAINNOT	30
5.5	PARAMETRIEN ESTIMOINTI LOGISTISESSA REGRESSIOSSA.....	31
5.6	NEWTON – RAPHSON METODI.....	32
6	TULOKSIA.....	35
6.1	ANALYYSISTRATEGIA	35

6.2	LOGISTISET REGRESSIOMALLIT KOLMEN VALITUN VAATERYHMÄN OSALTA	36
6.3	MUIDEN VAATERYHMIEN TULOKSIA	41
6.4	KORJATTU ALKUPERÄMAAJAKAUMA	41
7	JOHTOPÄÄTÖKSET JA POHDINTA.....	43
	LIITTEET	45
	LÄHTEET.....	49

1 JOHDANTO

1.1 Tutkielman taustaa

Tullihallituksen tilastoyksikkö kerää yrityksiltä tuonti- ja vientitietoja vuosittain julkaistavan määrityksen mukaisesti. Tilastoyksikkö julkaisee tilastoja Suomen kaupankäynnistä. Perusaineistona tutkielmassa on Suomen sisäkaupan vaateduonti vuodelta 2005.

Ulkomaankaupan tilastotiedot kerätään Euroopan unionissa (EU) kahdella eri järjestelmällä. Tilastotiedot kaupasta EU:n ulkopuolisten maiden kanssa (ulkokauppa) saadaan tullausjärjestelmästä. Tiedot jäsenmaiden välisestä kaupasta (sisäkauppa) kerätään erillisen menettelyn avulla, jota kutsutaan Intrastat-järjestelmäksi. (Intrastat, 2005)

Pro gradu -tutkielman keskeinen tavoite on selvittää, onko tullitilastoissa vääristyneisyyttä alkuperämaa merkinnän suhteen. Mikäli tutkielmasta ilmenee, että vääristymää on, pystyisivät käyttäjät huomioimaan sen olemassaolon tutkiessaan tilastoja.

1.2 Tiedonantovelvolliset tullitilastossa

Tullitilaston tiedonantovelvolliset ovat sisäkauppaa käyvät rekisteröidyt yritykset ja yhteisöt. Kukin EU:n jäsenmaa määrittelee kunkin vuoden lokakuussa viimeisen 12 kuukauden ajalta vuosituonnin ja -viennin arvoon perustuvan tilastointirajan. Tämä määritellään jäsenmaakohtaisesti siten, että mahdollisimman monet pienimmät tuojat ja viejät vapautuvat tilastoilmoituksen antamisvelvollisuudesta tilaston laatutason säilyessä kuitenkin riittävän hyvänä. (Intrastat, 2005)

1.3 Ulkomaankauppatilasto

Ulkomaankauppatilasto kuvaa Suomen ja muiden Euroopan unionin jäsenvaltioiden sekä Suomen ja Euroopan unionin ulkopuolisten maiden välistä tavarakauppaa eli sisä- ja ulkokauppaa. Ulkomaankauppatilasto on virallinen tietolähde Suomen tuonnista, viennistä ja kauppatasteesta. Euroopan yhteisön ulkokaupan (Extrastat) ja sisäkaupan (Intrastat) tilastointia koskevalla lainsäädännöllä varmistetaan, että tilastot perustuvat tarkoin määriteltyyn normistoon, jota sovelletaan kaikissa Euroopan unionin jäsenvaltioissa. Lisäksi ulkomaankaupan tilastoinnin käytännön soveltamisesta on annettu komission asetuksilla tai päätöksillä yhtenäiset määritelmät ja menetelmät. (Tulli, 2006a)

Tullihallitus tuottaa säännöllisesti useita erilaisia tilastoja. Julkaisukanava on pääasiassa tullin omat Internet-sivut. Osa julkaistaan myös painettuna, kuten vuosittainen Suomen ulkomaankaupan taskutilasto.

Yksityiskohtaiset tilastotiedot Suomen ulkomaankaupasta ovat myös saatavissa Internetin kautta sähköisestä ULTIKA-tietokannasta (<http://eportti.tietopalvelut.com>). Asiakkailta on pääsy tietokantaan ainoastaan maksusta. Tietokantaa päivitetään kerran kuukaudessa. (Tulli, 2006b)

1.4 Määrittelyä tuonnista ja viennistä

Pääsääntöisesti kaikki Suomesta vietävät ja Suomeen tuotavat tavarat sisältyvät sisä- ja ulkokaupan tavarakaupan tilastoon. Palvelujen kauppa jää tämän tilaston ulkopuolelle. Tilastoinnin edellytyksenä on, että tavara fyysisesti saapuu maahan tai lähtee maasta. Toisaalta tilastoinnin kannalta ei

ole merkitystä, sisältyykö tavaran tuontiin tai vientiin maksuliikennettä vai ei. (Tulli, 2006b)

1.5 Tutkimusongelmasta

Tutkielman aihe muotoutui tullin tarpeesta saada tietoa, onko Suomen vaatetuonnin alkuperämaajakauma vääristynyt sisäkaupan vaatetuonnin osalta. Tutkielmassa selvitetään mahdollisen vääristymän euromääräistä arvoa.

Tutkielmassa pyritään vastaamaan siihen, missä määrin alkuperämaa tieto on väärin kirjattu ja kuinka hyvin käytettävissä olevien tietojen perusteella on mahdollista estimoida luotettavasti Suomen sisäkaupan alkuperämaajakaumaa ja sen euroarvoa. Tullitilaston luotettavuuden kannalta olisi ensisijaisen tärkeää, että alkuperämaatieto olisi mahdollisimman oikein tilastoitu.

1.6 Miksi alkuperämaatieto on hyödyllinen

Suomessa alkuperämaatietoa on tilastoitu pitkään. Alkuperämaatiedon merkitys Suomen kaltaisessa maassa on tärkeä maantieteellisen sijaintinsa vuoksi, koska Suomeen tuodaan paljon kulutushyödykkeitä toisten maiden kautta. Esimerkiksi Suomen Pankilla on kiinnostusta alkuperämaatiedon laatuun, koska tiedon avulla maksutaseen arvo pystytään tilastoimaan oikealle alkuperämaatasolle tuonnissa ja viennissä.

Suomen maksutasetilaston laativat Tullihallitus, Tilastokeskus ja Suomen Pankki yhteistyössä. Tullihallitus vastaa tavarakaupan tilastoinnista, Tilastokeskus palvelukaupan ja tulonsiirtojen tilastoinnista ja Suomen

Pankki pääomakorvausten, pääomanliikkeiden sekä ulkomaisen varallisuusaseman tilastoinnista. (Suomen Pankki, 2005)

Ulkomaankauppatilasto on ensiarvoisen tärkeä väline lukuisille julkisen ja yksityisen sektorin päätöksentekijöille, suunnittelijoille ja tutkijoille. Seuraavassa esitetään muutamia ulkomaankauppatilaston käyttökohteita:

- Euroopan yhteisön viranomaisten monenkeskisten ja kahdenkeskisten kauppapolitiikkaa koskevien neuvottelujen valmistelu
- EU:n sisämarkkinoiden ja kansantalouden kehityksen seuranta
- muiden kansallisten ja kansainvälisten järjestöjen toimintapolitiikkojen valmistelu ja seuranta
- kansallisten intressien valvonta talous- ja kauppapolitiikassa
- yritysten tiedon lähde markkinatutkimuksien laadinnassa ja kaupallisten strategioiden kehittämisessä
- maksutasetilaston, kansantalouden tilinpidon ja taloudellisen tutkimuksen olennainen tietolähde
- globalisaation kehittymisen ja sen vaikutusten seuranta
- tiedotusvälineet ja "suuri yleisö"

Yllä esitetyt esimerkit kuvaavat ulkomaankauppatilaston käytön ja käyttäjäkunnan laajuutta ja monipuolisuutta. (Tulli, 2006a)

1.7 Tutkielman eteneminen ja menetelmät

Tutkielma on rajattu niin, että lähtömaana on aina Ruotsi ja alkuperämaina Ruotsi, Bangladesh, Indonesia, Intia, Kiina tai Pakistan. Seuraavassa kerrotaan tutkielman oletuksista, aineiston rajauksesta sekä logistisen regression käytöstä tutkielmassa. Aineiston käsittelyyn ja tilastollisiin ajoihin käytetään SAS ohjelman versiota 8.2.

Ruotsi valittiin lähtömaaksi siksi, että Ruotsi on Suomen tärkein kauppakumppani vaatetuonnissa. Edellä mainitut Aasian maat valittiin selittävän osan alkuperämaiksi, koska näissä maissa valmistetaan erittäin paljon vaatteita tuontiin Euroopan alueelle. Lisäksi Aasiassa vaatteiden tuotantokustannukset ovat matalammat ja siten vaatteiden hinta on selvästi halvempi kuin Euroopassa. Ruotsi jätettiin alkuperämaiden joukkoon, koska se osa, jossa Ruotsi on alkuperämaa, on tutkielmassa selitettävä osa. Asiaan perehdytään alla tarkemmin.

Vaihe 1: Datan muokkaus ja selittävät muuttujat

Datassa on 42 muuttujaa. Näistä poistetaan ne, jotka ovat merkityksettömiä. Merkityksettömät muuttujat ovat sellaisia, jotka eivät korreloi vastemuuttujan kanssa. Tämän jälkeen editoidaan Suomen tuontiaineistoa johtamalla aineistoon uusia muuttujia, *selittäviä muuttujia*, jotka korreloivat vastemuuttujan kanssa. Lisäksi dataan luodaan apumuuttujia, joilla ryhmitellään homogeeniset vaateryhmit.

Vaihe 2: Analyysisuunnitelma yleisellä tasolla

Tutkielmassa vastemuuttuja on binäärinen apumuuttuja, jota käyttämällä mallinnetaan logistiikkaketjua. Vastemuuttuja saa arvon 0,

jos alla olevista ehdoista kohta 1 toteutuu ja arvon 1, jos kohta 2 toteutuu.

1) Ruotsin tulliviranomaisten mukaan Ruotsissa ei juurikaan valmisteta vaatteita. Tämän vuoksi tutkielmassa oletetaan että, Ruotsissa ei valmisteta vaatteita tuontiin Suomeen. Eli alkuperämaatiedon ollessa Ruotsi, se on aina kirjattu väärin. Edellä kuvattu osa on rajattu tutkielmassa *selitettäväksi osaksi*.

2) Tutkielmassa rajattiin sisäkaupan aineistosta ne yritykset, jotka kirjaavat lähtömaatiedoksi Ruotsin ja alkuperämaatiedoksi Bangladesh, Indonesia, Intia, Kiina tai Pakistan. Tutkielmassa oletetaan, että yrityksen kirjatessa näin, se on aina kirjattu oikein. Edellä kuvattu osa on tutkielmassa *selittävä osa*.

3) Selittävälle osalle aineistoa, joka saatiin kohdasta 2, sovitetaan logistinen regressiomalli. Tässä mallinnetaan numeron yksi todennäköisyyttä suhteessa arvoon nolla. Saadaan tilastollinen malli logistiikkaketjusta, joka tutkielmassa oletetaan olevan likimain oikea.

4) Käytetään kohdasta 3 saadun mallin parametrikertoimia selitettävään osaan 1 niin, että tuloksena saadaan jokaiselle erälle todennäköisyysarvo kuulua mallinnettuun logistiikkaketjuun. Mikäli todennäköisyysarvo jollakin vaate-erällä on vähintään 50 %, korvataan selitettävän kohdan yksi alkuperämaatieto uudella tiedolla AA, joka kuvaa maita Bangladesh, Indonesia, Intia, Kiina ja Pakistan. Oletuksena on, että estimoitu malli tuottaisi luotettavamman alkuperämaajakauman vaatetuonnista.

Vaihe 3: Uusi alkuperämaajakauma

Vaiheessa 3 lasketaan alkuperämaatiedon AA euromääräinen kokonaisarvo eri vaateryhmissä ja saadaan uusi alkuperämaajakauma. Kiinnostuksen kohteena tutkielmassa on se osa, joka on oletuksien mukaan väärin kirjattu alkuperämaatasolla.

Alla oleva kuva havainnollistaa selittävää ja selitettävää logistiikkaketjua. Rivillä yksi on kuvattuna selittävä osa ja rivillä kaksi selitettävä. Rivin yksi ehdoilla vaste saa arvon 1 ja rivin 2 ehdoilla 0.

alkuperämaa		lähtömaa		kohdema
1) Aasia	→	Ruotsi	→	Suomi
2) Ruotsi	→	Ruotsi	→	Suomi

Kuvio 1. Logistiikkaketjut

1.8 Euroopan unionin sisäkaupan perusasetus

1. Yrityksillä ei ole ilmoitusvelvollisuutta alkuperämaan suhteen, koska Euroopan unionin sisäkaupan perusasetus ei tätä tietoa vaadi. Seuraavassa on selvitetty tarkemmin miksi alkuperämaatieto voi olla väärin kirjattu.

- Sähköistä tullilmoitusta käytettäessä alkuperämaatieto on pakollinen, mutta ei ole tiedossa onko tieto kirjattu oikein. Tietoteknisistä syistä ei ole olemassa tietoa, missä tullilmoituksissa alkuperämaatieto on puuttunut tai muutettu.
- Mikäli tullilmoitus on tehty lomakkeella ja alkuperämaatieto puuttuu, tullin työntekijä kirjaa alkuperämaaksi lähtömaan.

2. Yritysten on helppo kirjata lähtömaa alkuperämaaksi, koska alkuperämaatiieto ei ole pakollinen. Otetaan esimerkiksi, että Kiinasta on tuotu laivalastillinen vaatteita. Tämä laivalasti merkitään Euroopan unionin rajan ohitettuaan Kiinasta tuoduksi. Oltuaan jonkin aikaa välivarastossa alkuperämaatiettoa ei välttämättä enää löydy, koska tieto on voinut kadota esimerkiksi vaate-erien omistajavaihdoksien yhteydessä. Vaihtoehtoisesti syynä voi esimerkiksi olla myös yritysten tietojärjestelmien puutteellisuus tai alkuperämaatiettoa on vaikea saada selville.

1.9 Tutkielman tavoitteet tiivistetysti

Tutkielmassa mallinnetaan, mitkä tavaraerät on valmistettu Aasiassa, vaikka tilastoilmoituksessa on alkuperämaaksi kirjattu Ruotsi. Selittävään osan jokaiselle vaateryhmälle sovitetaan logistinen regressiomalli, josta saadaan selittäville muuttujille parametrikertoimet. Edellä mainittuja tilastollisia malleja käytetään mallintamaan tutkielman selitettävää aineisto-osaa, joka määritellään luvussa 3.6.

Pro gradu -tutkielman keskeisin tavoite on tehdä soveltuvien tilastollisten tarkastelujen avulla arvio siitä, kuinka suuressa osassa (euromääräinen arvo) vaatetuonnista alkuperämaatiieto on väärin ilmoitettu. Tämä osuus pyritään mallintamaan oikealle alkuperämaatasolle. Tilastollisella mallilla jokaiselle vaate-erälle saadaan todennäköisyysarvo kuulua mallinnettuun logistiikkaketjuun. Tullitilastoissa tutkittavan osan alkuperämaajakauma on euromääräisesti jakautunut alla olevan taulukon 1 mukaisesti.

Taulukko 1. Euromääräinen alkuperämaajakauma

maa	summa euroa	erät	prosenttia / kokonaisarvosta
SE	79 540 487	23008	45,00 %
CN	60 588 465	60307	34,45 %
BD	17 581 772	13253	10,00 %
IN	11 917 238	23443	6,70 %
PK	3 991 479	6975	2,27 %
ID	2 178 802	2805	1,24 %

Tutkielmassa tutkitaan alkuperämaatiedon laatua ja sen avulla pyritään vastaamaan seuraaviin kysymyksiin:

- 1) Onko alkuperämaajakauma väärin jakautunut.
- 2) Paljonko on mahdollisen vääristymän euromääräinen arvo.
- 3) Miten saadun laadullisen arvion perusteella olisi mahdollista kehittää tulevaisuudessa alkuperämaatietojen keräämistä ja tilastointia.

2 KESKEISET KÄSITTEET

Tässä luvussa määritellään pro gradu -tutkielmani kannalta keskeiset käsitteet. Keskeisimmät käsitteet työni kannalta ovat alkuperämaa, lähtömaa, sisäkauppa ja tilastoarvo.

Alkuperämaa: Maa, jossa tavara on tuotettu tai valmistettu. Mikäli tavaraa on valmistettu kahdessa tai useammassa maassa, alkuperämaa on maa, jossa viimeisin merkittävä ja taloudellisesti perusteltu valmistus tai käsittely on tapahtunut. Tavarankäsitteeksi tullin työntekijä kirjaa lähtömaan silloin, kun alkuperämaa ei ole tiedossa. (Intrastat, 2005)

Lähtömaa: Tuonnin lähtömaalla tarkoitetaan EU:n jäsenmaata, josta tavara on alun perin lähetetty vietäväksi Suomeen joko suoraan tai toisen maan kautta. (Intrastat, 2005)

Sisäkauppa: Kauppaa, jota Suomi harjoittaa jonkun Euroopan unionin maan kanssa. (Intrastat, 2005)

Tilastoarvo: Tuonnissa arvonlisäveron laskemisen pohjana oleva verotusarvo, josta on vähennettävä verotusarvoon kuuluva mahdollinen valmistevero silloin, kun se kannetaan Suomeen tuonnin yhteydessä. Tilastoarvoon on sisällytettävä rahti- ja vakuutuskustannukset Suomen rajalla olevaan ensimmäiseen rajanylityspaikkaan saakka. (Intrastat, 2005)

Ulkokauppa: Kauppaa, jota Suomi harjoittaa maan kanssa, joka ei kuulu Euroopan unioniin. (Intrastat, 2005)

Tiedonantovelvollinen: On yritys tai yhteisö, josta merkitään tiedonantajan (tuoja tai viejä) täydellinen nimi, osoite ja yritystunnus. Tunnuksena käytetään yrityksen alv-numeroa. (Intrastat, 2005)

Tilastointiraja: Tilastotietojen ilmoittamisen kynnyсарvo oli vuonna 2005 tuonnissa 100 000 euroa ja viennissä 200 000 euroa. Tilastotiedot voi ilmoittaa joko lomakkeella tai sähköistä tietojenvälitystapaa käyttäen. (Intrastat, 2005)

Sähköinen ilmoitus: Sähköistä ilmoitusta ei voi lähettää ellei alkuperämaa tietoa ole saatavilla. Lomakkeella täytettäessä tullin työntekijä korvaa alkuperämaatiedon lähtömaalla, mikäli se puuttuu.

Tulli-ilmoitus: Tulli-ilmoituksella ilmoitetaan tullattavien tavaroiden tiedot. Esimerkki tulli-ilmoituksesta löytyy liitteestä 2.

Aasia: Tutkielmassa Aasialla tarkoitetaan seuraavaa viittä maata: Bangladesh, Indonesia, Intia, Kiina ja Pakistan. Nämä viisi maata valittiin, koska tullin mukaan nämä maat ovat samankaltaisia Kiinan kanssa tuonnissa Euroopan alueelle.

Jälleentuonti: Euroopan rajan ylittäessään tuotteesta maksetaan tullimaksu. Tämän jälkeen tavarankulkeutumista Euroopan rajojen sisällä on lähes mahdotonta seurata. Tätä kulkeutumista rajojen sisällä kutsutaan jälleentuonniksi.

Logistiikkaketju: Tutkielmassa logistiikkaketjulla tarkoitetaan tavaroiden kuljetusta alkuperämaasta kohdemaahan. Tavarointa voidaan kuljettaa useiden maiden kautta.

3 SUOMEN SISÄ- JA ULKOKAUPPA

3.1 Suomen sisäkauppa-aineisto (Intrastat)

Sisäkaupan vaatetuonnin kokonaisarvo vuonna 2005 Suomeen oli noin 800 miljoonaa euroa. Ruotsista Suomeen tuotiin vaatteita 233 miljoonalla eurolla, mikä oli 29 % koko sisäkaupan vaatetuonnista.

3.2 Suomen ulkokauppa-aineisto (Extrastat)

Ulkokaupan aineistossa on tavaraeriä 43 630 ja yrityksiä on 2685. Ulkokaupan vaatetuonnin koko arvo oli 282 miljoonaa euroa vuonna 2005. Yrityksiä, jotka tuovat vaatteita sekä ulkokaupan että sisäkaupan alueelta on 377 yritystä. Kiina oli suurin vaatetuojia ulkokaupassa vuonna 2005.

3.3 Eurostatin aineisto koko Euroopan unionin alueelle

Eurostatin tietovarastot sisältävät karkealla tasolla kaiken viennin ja tuonnin koko Euroopan unionin alueelle ja alueella. Eurostatin Internet-sivuilla on HTML portaali pitkältä aikaväliltä kerättyihin tilastoihin. Tämä aineisto on vapaasti saatavilla. Paremman kokonaiskuvan vuoksi tutkielmaan on lisätty lukuja Eurostatin tietokannasta. Eurostatin tietovarastoista saatu vaatetuonnin euromääräinen arvo koko Euroopan unionin alueelle Euroopan unionin ulkopuolelta vuonna 2005 oli 35 miljardia euroa.

4 AINEISTON RAJAUS

Aineistona on Suomen tuontitilasto vuodelta 2005. Tutkielmassa ollaan kiinnostuneita vaatetuonnista, koska siinä erityisesti on havaittu olevan epävarmuutta alkuperämaan osalta. Tutkielmassa lähtömaaksi valittiin Ruotsi, koska Ruotsista tuodaan Suomeen euromääräisesti eniten vaatteita. Tutkimuksessa käytettäväksi aineistoksi muotoutui matriisi, jossa on 129 791 riviä (tuontierää) ja 16 saraketietoa.

Tulli saa tuonti- ja vientiaineiston seuraavalla tavalla: Yrityksen ylittäessä määritellyt tuonti- tai vientirajat tulee näistä tiedonantovelvollisia. Tilastoyksikössä näistä yritystiedoista ja verottajalta saatavista tiedoista kootaan tuonti – ja vientiaineisto. Tärkeimmät aineiston muuttujien määritelmät löytyvät liitteestä 1.

4.1 Tutkimusaineisto ja aineiston rakenne

Aineistoon luotiin johdettuja selittäviä muuttujia. Näitä muuttujia käytetään työssä mallintamaan tutkittavaa logistiikkaketjua. Alla kuvataan nämä selittävät muuttujat.

Yksikköarvo (yarvo) = Jatkuva muuttuja. Erän tilastoarvo jaettuna painolla. Saadaan vaatteen arvo kilogrammalta.

Kuljetusmuoto (kulj) = Indikaattorimuuttuja, saa arvon 1 jos laiva, 0 muulloin.

Paino = Jatkuva muuttuja. Kuinka paljon erä painaa (kg).

Sesonki = Indikaattorimuuttuja, joka saa arvon 1, jos tammi-, helmi-, maaliskuu-, loka-, marras- tai joulukuu, 0 muulloin.

Aineiston rakennetta on kuvattu taulukossa 2. Taulukossa on mukana tutkielman kannalta keskeisiä sarakkeita. Aineistoon luotiin myös muuttujia, joilla yhdistettiin homogeenisiä tavararyhmiä. Tavararyhmät on kuvattu liitteessä 1, taulukosta 2 nämä sarakkeet puuttuvat.

Taulukko 2. kuvaus aineiston rakenteesta

tavaraerä	yritys	alkuperämaa	lähtömaa	arvo	kulj	paino	sesonki
1.	1	SE	SE	100	0	1	0
2.	1	SE	SE	15000	1	20	0
3.	1	ID	SE	20	1	1	1
4.	2	PK	SE	50000	1	2000	1
5.	3	BD	SE	150	1	14	0
6.	4	CN	SE	4200	1	12	1
.
129791	M	maa	lmaa	176milj.	.	.	.

4.2 Vaateryhmien yhdistäminen

Aineistossa yksi rivitieto vastaa yhtä tavaraerää, siis tutkielmassa yhtä vaate-erää. CN-nimikkeistö (Combined Nomenclature) on Euroopan yhteisön yhdistetty nimikkeistö, jonka 8-numeroisia nimikkeitä käytetään vienti-ilmoituksissa ja sisäkaupan tilastoilmoituksissa. Nimikkeistöön tulee muutoksia vuosittain ja seuraavan vuoden alussa voimaan tuleva nimikkeistö julkaistaan joka vuosi viimeistään lokakuun loppuun mennessä yhteisön virallisessa lehdessä (Tulli, 2007). Nimikkeistön avulla pystytään yksiselitteisesti määrittämään millaisesta tavarasta on kyse. Seuraavassa kuvataan miten aineistoon johdettiin homogeeniset vaateryhmät.

Aineistossa on sarake, joka määrittää tavaraerän sisällön kahdeksanmerkkisenä CN8-nimikkeistön avulla. Sarake muokattiin aluksi neljämerkkiseksi CN4-tasolle, jonka avulla on helpompi yleisemmällä tasolla erottaa tavaraerän sisältö. CN4-tasosta vaatteet ryhmiteltiin kahdeksaan eri homogeeniseen vaateryhmään, joita ovat: Takit, puvut, paidat, vauvanvaatteet, alusvaatteet, urheiluvaatteet, muut vaatteet ja muut tekstiilit. Vaateryhmien tarkemmat CN4 tasot ja selitykset liitteessä 1.

Taulukosta 3 ilmenee se, että suurin osa Ruotsin kautta Suomeen tuotavista vaatteista valmistetaan Aasiassa. Tälle osalle tutkielmassa mallinnetaan logistinen regressiomalli.

Taulukko 3. Suomeen tuotujen vaate-erien määrät ja euroarvot alkuperämaittain lähtömaan ollessa Ruotsi (2005). Taulukossa on esitetty kuusi suurinta erien tuojaa.

Alkuperämaa	erien määrä	erien yhteisarvo euroa
1. Kiina	60307	60 588 465e
2. Intia	23443	11 917 238e
3. Bangladesh	13253	17 581 772e
4. Hongkong	11564	6 960 047e
5. Turkki	10970	14 604 030e
6. Pakistan	6975	3 991 479e
..	.	.

4.3 Selittävä osa aineistosta

Selittävä osa viittaa aineistoon, jossa alkuperämaatiedon oletetaan olevan oikein kirjattu. Taulukko 4 sisältää selittävän osan erien määrät ja niiden arvot. Taulukossa on mukana alkuperämaina valitut Aasian maat ja lähtömaana on Ruotsi.

Taulukko 4. Selittävä osa

ryhmä	yriytysten määrä	erien määrä	arvo (milj. EUR)
selittävä osa	237	106 783	96.3

Selittävän aineiston muokkaus:

Aineistoon luotiin uusia muuttujia, joilla pyritään selittämään tutkittavaa kokonaisilmiötä. Tilastollinen malli tehdään jokaiselle vaateryhmälle erikseen.

Kokonaisaineistosta muokattiin selittävä aineisto seuraavalla tavalla:

1) Luotiin uudet binääriset muuttujat a_1, a_2, \dots, a_8 , jotka toteuttavat liitteessä 1 määritellyt ehdot vaateryhmille. Muuttuja a_1 on ryhmä takit, a_2 puvut jne.

2) Aineistoon luotiin binäärinen vastemuuttuja. Vastemuuttuja saa arvon 1, jos alkuperämaa on Bangladesh, Indonesia, Intia, Kiina tai Pakistan ja lähtömaa on Ruotsi, muuten vastemuuttuja saa arvon 0. Jokaiselle selittävän osan homogeeniselle vaateryhmälle laskettiin logistiset regressiomallit erikseen.

Edellä olevat alkuperämaat ryhmiteltiin yhteen, koska tutkielmassa oletamme niiden olevan samankaltaisia Kiinaan nähden vaatetuonnissa Suomeen.

Taulukossa 5 on kuvattuna jatkuvien selittävien muuttujien tunnuslukuja. Vaihteluvälit erien yksikköarvoissa ja painoissa ovat varsin suuria.

Taulukko 5. Jatkuvien selittävien muuttujien tunnuslukuja

Muuttuja	N	Mean	Std Dev	Minimum	Maximum
yksikköarvo	106 783	16.5	40.6	0.0	35836.0
paino (kg)	106 783	47.0	344.8	1.0	19677.0

Taulukosta 6 nähdään, että yritykset toivat Suomeen suhteellisen edullisia eriä. 0-40 euroa/kg oli koko datasta oli lähes 50 %.

Taulukko 6. Yksikköarvojen frekvenssitaulukko / yritys

yksikköarvo	luokka	frekvenssi	prosenttiosuus	kum. prosentti
0-10	1	24632	23.07	23.07
10-20	2	13552	12.69	35.76
20-30	3	8322	7.79	43.55
30-40	4	5934	5.56	49.11
40-50	5	4413	4.13	53.24
50-60	6	3605	3.38	56.62
60-70	7	2902	2.72	59.34
70-80	8	2558	2.40	61.73
80-90	9	2129	1.99	63.72
90-	10	38736	36.28	100.00

Taulukko 7 esittää erien painon frekvenssitaulukon. Taulukosta nähdään, että erät olivat pääosin kevyitä. Suomessa kaukomyyntiä harjoittavan ulkomaisen yrityksen on hakeuduttava täällä verovelvolliseksi, mikäli yksityishenkilölle tapahtuvan myynnin arvo on yli 35 000 euroa vuodessa (Intrastat, 2005). Kevyet erät selittyvät sillä, että postimyynti ja Internet-ostokset ulkomailta merkitään tullitilastoihin tuonniksi edellä olevan ehdon toteutuessa.

Taulukko 7. Painojen (kg) frekvenssitaulukko

erän paino (kg)	luokka	frekvenssi	prosenttiosuus	kum. prosentti
< 1	1	39401	36.90	36.90
1-4	2	22225	20.81	57.71
4-12	3	18171	17.02	74.73
12-50	4	17111	16.02	90.75
>50	5	9875	9.25	100.00

Taulukossa 8 on esitetty selittävien muuttujien välinen korrelaatiomatriisi. Taulukosta nähdään, että korrelaatiot ovat varsin heikkoja. Kaikkia muuttujia voidaan käyttää selittäjinä tilastollisissa malleissa.

Taulukko 8. Selittävien muuttujien välinen korrelaatiomatriisi

	yksikköarvo	kuljetusmuoto	sesonki	paino (kg)
yksikköarvo	1.00000	-0.04145	0.00335	0.00884
		<.0001	0.2733	0.0039
kuljetusmuoto	-0.04145	1.0000	0.00599	-0.02970
	<.0001		0.8722	<.0001
sesonki	-0.0009	-0.0003	1.0000	-0.0057
	0.6308	0.8722		0.0021
paino (kg)	-0.0254	-0.0158	-0.0057	1.0000
	<.0001	<.0001	0.0021	

4.4 Selitettävä osa aineistosta

Selitettävä osa viittaa aineistoon, jossa alkuperämaatiedon oletetaan olevan väärin kirjattu. Taulukko 9 sisältää selitettävän osan yritysten ja erien määrät sekä niiden euroarvot. Taulukossa alkuperä- ja lähtömaana Ruotsi.

Taulukko 9. Selitettävä osa

ryhmä	yriytysten määrä	erien määrä	arvo (milj. EUR)
selitettävä osa	852	23 008	79.5

Seuraavassa kuvataan tutkielman selitettävää osaa. Selitettävällä osalla tarkoitetaan aineiston osaa, joka oletetaan väärin merkityksi alkuperämaatasolla.

Selitettävän aineiston muokkaus:

Kokonaisaineistosta muokattiin selitettävä aineisto seuraavalla tavalla:

1) Luotiin uudet binääriset muuttujat a_1, a_2, \dots, a_8 , jotka toteuttavat liitteessä 1 määritellyt ehdot vaateryhmillä. Muuttuja a_1 on ryhmä takit, a_2 puvut jne.

2) Aineistoon luotiin vastemuuttuja, joka on binäärinen. Vastemuuttuja saa arvon 1, jos alkuperämaina on joki valituista Aasian maista ja lähtömaa on Ruotsi, muuten vastemuuttuja saa arvon 0. Eli selitettävässä osassa vastemuuttuja saa arvon 0 aina.

Tutkielmassa koko selitettävän osan arvo oli 79,5 miljoonaa euroa, tämä oli Suomen vaatetuonnista noin 10 % vuonna 2005.

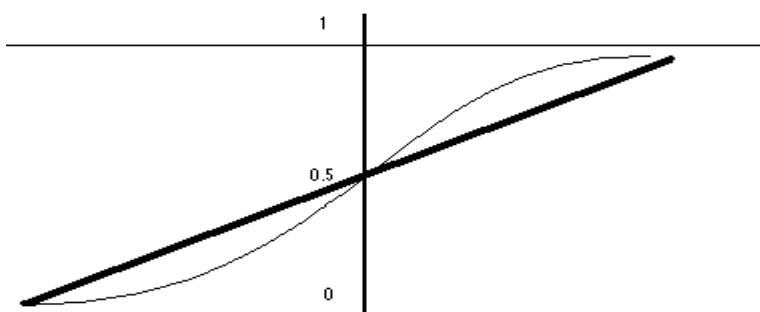
5 TILASTOLLISET MENETELMÄT

5.1 Logistinen regressio

Aineistoon muodostettiin uusi binäärinen muuttuja y , joka on tutkielmassa vastemuuttuja. Logistisen regression erottaa lineaarisesta regressiosta se, että logistisessa regressiossa vastemuuttuja on binäärinen. Tämä ero vaikuttaa sekä mallin valintaan että oletuksiin. Riippumattomat eli selittävät muuttujat $x = (x_1, x_2, \dots, x_p)$ voivat olla jatkuvia, luokiteltuja tai järjestysasteikollisia. (Hosmer & Lemeshow, 1989, s. 1).

$$y_k = \begin{cases} 1, & \text{jos vaate-erä } k \text{ täyttää luvussa 4.3 vaaditut ehdot} \\ 0, & \text{jos vaate-erä } k \text{ ei täytä luvussa 4.3 vaadittuja ehtoja} \end{cases}$$

Lineaarisisessa mallissa jossa, vastemuuttuja on jatkuvatyypinen, selittäjien ja vasteen yhteys on lineaarista muotoa (paksu suora viiva, kuvio2). Logit-malleissa vastemuuttujan odotettujen arvojen epälineaarinen funktio (s:n muotoinen viiva) on yhteydessä selittäjiin lineaarisesti.



Kuvio 2. logistinen ja lineaarinen funktio graafisesti

5.2 Selitysasteen määritelmä logistisessa regressiossa

Selitysaste, jota SAS ohjelma käyttää logistisessa regressiossa on Cox & Snell:n määrittelemä vuodelta 1989 ja se on muotoa:

$$R^2 = 1 - \left\{ \frac{L(0)}{L(\hat{\theta})} \right\}^{\frac{2}{n}}$$

Missä $L(0)$ on logaritminen uskottavuus mallille, jossa parametrina on ainoastaan vakio-termi. $L(\hat{\theta})$ on taas koko mallin logaritminen uskottavuus. (Cox & Snell, 1989, s. 208-209)

$$R_{\max}^2 = 1 - \{L(0)\}^{[2/n]}$$

Uudelleenskaalattu maksimiselitysasteen määritelmä, jota SAS ohjelma käyttää on Nagelkerken määrittelemä vuodelta 1991 ja se on muotoa:

$$\tilde{R}^2 = \frac{R^2}{R_{\max}^2}$$

Tässä \tilde{R}^2 on uudelleenskaalattu maksimiselitysaste. Nagelkerken selitysaste ottaa huomioon sen, että logistisella mallilla ei teoreettisestikaan voida päästä 100 %:n selitykseen. (Nagelkerke, 1991, s. 691-692)

5.3 Logistinen regressio, riippumattomat havainnot

Tutkielmassa ilmiön tutkimiseen käytetään logistista regressiota ja logit muunnosta. Selittävää osaa mallinnettiin niin, että vastemuuttuja y sai arvon 1 tai 0, jokaiselle vaate-erälle. Seuraavassa tarkastelussa oletetaan, että vastemuuttuja on binäärinen.

$$y_k = \begin{cases} 1, & \text{jos tuontierä } k \text{ täyttää luvussa 4.3 vaaditut ehdot} \\ 0, & \text{jos tuontierä } k \text{ ei täytä luvussa 4.3 vaadittuja ehtoja} \end{cases}$$
$$k = 1, 2, 3, \dots, N$$

Tämän lisäksi oletetaan, että $\pi(y_k = 1) = \pi$ ja $\pi(y_k = 0) = 1 - \pi$, tällöin

$y_k \sim Be(1, \pi)$ ja Be tarkoittaa Bernoulli jakaumaa ja tässä,

$$E(y_k) = \mu_k = \pi \text{ ja } \text{var}(y_k) = \pi(1 - \pi).$$

Logistinen funktio on muotoa:

$$\pi(\mathbf{x}) = \frac{e^{g(\mathbf{x})}}{1 + e^{g(\mathbf{x})}}$$

(Hosmer & Lemeshow, 2000, s. 32)

Tässä $\pi(\mathbf{x})$ saa arvoja väliltä $[0, 1]$ siten, että $\pi(-\infty) = 0$ ja $\pi(\infty) = 1$. Funktiosta saatavaa arvoa voidaan ajatella todennäköisyytenä sille, kuinka todennäköisesti satunnainen erä k on tullut tutkielmassa mallinnettua logistiikkaketjua pitkin Suomeen.

Logistisessa funktiossa $g(\mathbf{x})$ lineaarinen lauseke on seuraavanlainen:

$$g(\mathbf{x}) = \beta_0 + \beta_1 x_{1k} + \beta_2 x_{2k} \dots + \beta_p x_{pk},$$

(Hosmer & Lemeshow, 2000, s. 31)

Selittäjinä toimivat x_1, \dots, x_p , jotka voivat olla joko luokiteltuja tai jatkuvia. Regressiokertoimia ovat $\beta_0, \beta_1, \beta_2, \dots, \beta_p$, joista β_0 on vakiotermi ja $\beta_1, \beta_2, \dots, \beta_p$ ovat regressiokertoimia, joista jokainen vaikuttaa selittävien muuttujien kautta funktion arvoon.

Logistinen funktio linearisoituu logit-muunnoksella:

$$\text{logit}(\pi(y_k = 1 | \mathbf{x}_k)) = \log\left(\frac{\pi(\mathbf{x})}{1 - \pi(\mathbf{x})}\right) = \mathbf{x}'_k \boldsymbol{\beta},$$

missä

$$\mathbf{x}_k = (1, x_{1k}, \dots, x_{pk})' \text{ ja}$$

$$\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_p)'$$

5.4 Logistinen regressio, riippuvat havainnot

Oletetaan, että X_i on otos toisistaan riippuvia havaintoja rypäissä. Tässä voimme käyttää GEE (generalized estimating equations) metodologiaa. GEE-menetelmä on selostettu tarkemmin julkaisussa Longitudinal data analysis using generalized linear models, Liang & Zeger 1986. GEE voi tarvittaessa ottaa huomioon vaihtuvan korrelaattorakenteen ja havainnot voivat riippua toisistaan rypäiden sisällä.

Havaintojen kovarianssirakenne rypäiden sisällä voidaan esittää muodossa.

$$V_i = \phi A_i^{\frac{1}{2}} R(\alpha) A_i^{\frac{1}{2}} \quad i=1, \dots, m$$

Tässä A_i on varianssien $V(y_k)$ diagonaalimatriisi rypäessä i . $R(\alpha)$ on korrelaatiomatriisi ennaltamääritellystä korrelaatioparametrien α havainnoista rypäessä i . Parametri ϕ on hajontaparametri. Vaihtuvan sisäkorrelaatorakenteen tapauksessa parametri α on skalaari ja se on estimoitava. Yleensä estimointi menetelmänä \hat{b} laskemiseksi on käytetty Newton-Raphson iterointialgoritmia. (Lehtonen & Pahkinen, 2004, s. 287-288)

5.5 Parametrien estimointi logistisessa regressiossa

Tarkastellaan yksinkertaisuuden vuoksi tilannetta, jossa on yksi selittävä muuttuja x . Oletamme havainnot riippumattomiksi.

Saamme uskottavuusfunktion, joka on muotoa:

$$l(\beta) = \prod_{k=1}^n p(x_k)^{y_k} (1 - p(x_k))^{1-y_k}$$

josta saamme logaritmisen uskottavuusfunktion, joka on seuraavaa muotoa:

$$L(\beta) = \log[l(b)] = \sum_{i=1}^n \{y_k \log[p(x_k)] + (1 - y_k) \log[1 - p(x_k)]\}$$

Muunnos logaritmiseksi tehtiin, jotta suurimman uskottavuuden estimaatti olisi helpompi laskea matemaattisesti. Uskottavuusfunktio $L(\beta)$ tulee derivoida β_0 ja β_1 suhteen ja asettaa näin saatu funktion arvo nolllaksi. Saadaan seuraavaa:

$$\sum_{k=1}^n [y_k - p(x_k)] = 0$$

Ratkaisuna yhtälöstä saadaan suurimman uskottavuuden estimaatti, jonka merkintätapa on $\hat{\beta}$. Logistisessa regressiossa funktion lauseke on epälineaarinen ja sen vuoksi ratkaisussa tarvitsee käyttää iteratiivisia menetelmiä. (Hosmer & Lemeshow, 2000, s. 8-9)

5.6 Newton – Raphson metodi

Tutkielmassa suurimman uskottavuuden selvittämisessä käytettiin SAS ohjelmaa ja sen logistic proseduuria, joka käyttää Newton – Raphson - iterointialgoritmia laskiessaan estimaatin arvoa. Kun suurimman uskottavuuden estimaattia ei voida ratkaista suljetussa muodossa, joudutaan käyttämään iterointimenetelmiä. Ajatuksena on kehittää uskottavuusyhtälö U Taylorin sarjaksi parametrivektorin β alkuarvovektorin β_0 suhteen. Alkuarvovektori β_0 on arvaus, josta iterointi aloitetaan. Yhtälö on seuraavanlainen:

$$U(\beta) = \frac{\partial l(\beta)}{\partial \beta} = 0$$

Taylorin sarjakehitelmä yleiselle funktiolle on seuraavanlainen.

$$f(x) = f(a) + (x-a)f'(a) + f''(a)\left[\frac{(x-a)^2}{2!}\right] + \dots + R_n$$

$R_n = \text{Residuaali}$, $R_n \xrightarrow{n \rightarrow \infty} 0$

Kehitetään uskottavuusyhtälö U Taylorin sarjaksi. Käytetään linearisointimenetelmää ja tehdään seuraava approksimaatio.

$$U(\beta) \approx U(\beta_0) + (\beta - \beta_0)U'(\beta_0)$$

Tällöin

$$U(\hat{\beta}) \approx U(\beta_0) + (\hat{\beta} - \beta_0)U'(\beta_0) = 0, \text{ jos } \hat{\beta} \approx \beta_0$$

tästä seuraa

$$U(\beta_0) + \hat{\beta}U'(\beta_0) = \beta_0U'(\beta_0)$$

$$\Rightarrow \hat{\beta} = \frac{\beta_0U'(\beta_0) - U(\beta_0)}{U'(\beta_0)}$$

$$\Rightarrow \hat{\beta} = \beta_0 - \frac{U(\beta_0)}{U'(\beta_0)}$$

Tässä uusi approksimaatio suurimman uskottavuuden yhtälöiden on saatu $\hat{\beta}$. Olkoon tämä uusi ratkaisu $\hat{\beta}_k$. Seuraavaksi ratkaistaan uusi approksimaatio $\hat{\beta}_{k+1}$.

$$\hat{\beta}_{k+1} = \hat{\beta}_k - \frac{U(\hat{\beta}_k)}{U'(\hat{\beta}_k)} \quad k=0,1,2,3\dots$$

Tästä saadaan jatkamalla yleinen kaava Newton-Raphson iterointialgoritmille. Iterointikierröksiä tulee tehdä niin monta, että yhtälö konvergoituu lähelle arvoa $\hat{\beta}_{k+1}$ niin, että suurimman uskottavuuden estimaatiksi tulee $\hat{\beta} = \hat{\beta}_{k+1} \Rightarrow \beta$. (Ortega & Rheinboldt, 2000, s. 181-183)

6 TULOKSIA

Tutkielman tuloksiin valittiin tarkemmin analysoitavaksi kolme vaateryhmää. Ryhmät valittiin, koska niissä alkuperämaajakaumat muuttuivat eniten euromääräisesti laskettuna. Tuloksissa esitetään euroarvot kaikkien vaateryhmien osalta ennen mallien sovittamista selitettävään osaan sekä sovittamisen jälkeen.

Alla esitetään selittävät muuttujat ja niiden skaalaus.

x_1 = yksikköarvo	jatkuva
x_2 = kuljetusmuoto	0/1
x_3 = sesonki	0/1
x_4 =paino	jatkuva

(katso sivu 22, kovariaatit)

6.1 Analyysistrategia

1) Tutkielmassa sovitetaan logistinen regressiomalli aineiston selittävälle osalle, jonka oletetaan olevan alkuperämaatasolla oikein kirjattu. Selittävä osa aineistoa jaettiin kahdeksaan osaan vaateryhmittäin ja jokaiselle vaateryhmälle laskettiin logistinen regressiomalli, kuten luvussa 3.6 on esitetty.

Logistinen regressiomalli on muotoa:

$$\text{logit}(\pi(y_k = 1 | \mathbf{x}_k)) = \log\left(\frac{\pi(\mathbf{x})}{1 - \pi(\mathbf{x})}\right) = \mathbf{x}'_k \boldsymbol{\beta}$$

Saadaan estimoitu $\hat{\beta}$ vektori malleille.

2) Lasketaan saadulla logistisella mallilla ennusteet selitettävälle osalle, joka oletetaan väärin merkityksi, kuten luvussa 3.7 on esitetty.

Ennusteet ovat muotoa

$$\hat{y}_k = \mathbf{x}'_k \hat{\beta}$$

3) Saadaan ennustetut \hat{y}_k todennäköisyydet kaikille vaate-erille 1,...,n. Tässä todennäköisyysarvo \hat{y}_k kuvaa sitä, että logistiikkaketju on samanlainen kuin selittävässä osassa. Eli alkuperämaa olisi Ruotsin sijasta Bangladesh, Indonesia, Intia, Kiina tai Pakistan.

6.2 Logistiset regressiomallit kolmen valitun vaateryhmän osalta

Pukujen tuontia mallinnettiin logistisella regressiomallilla. Aineistosta rajattiin alkuperämaiksi Aasia, lähtömaaksi Ruotsi ja vaateryhmäksi puvut.

Logistisen regressiomallin tilastollisesti merkitsevät muuttujat ja niiden interaktiot on kuvattu alla. Mallin parametriestimaatit ja tilastollisia tunnuslukuja on esitetty taulukossa 10.

logit P(Y)=vakio+yarvo+kulj+sesonki+yarvo*sesonki+paino+yarvo*paino+kulj*paino

Taulukko 10. Mallin parametriestimaatit, kun alkuperämaana Aasia ja ryhmänä puvut

	Analysis	of Maximum	Likelihood	Estimates	
parametric	DF	estimaatti	SE	Chi-Square	p - ChiSq
Intercept	1	-0.7135	0.0835	73.1015	<.0001
yarvo	1	-0.00400	0.000548	53.1665	<.0001
kulj	1	0.0184	0.0826	0.0495	0.8240
sesonki	1	-0.1612	0.0178	82.2772	<.0001
yarvo*sesonki	1	0.00199	0.000735	7.3457	0.0067
paino	1	-0.00462	0.00103	20.2470	<.0001
yarvo*paino	1	0.000015	1.718E-6	78.6843	<.0001
kulj*paino	1	0.00410	0.00103	15.9344	<.0001

Taulukko 11. Mallin selitysasteet vaateryhmissä puvut

R-Square	0.0033	Max-rescaled	R-Square	0.0047
----------	--------	--------------	----------	--------

Lopullisessa mallissa puvuille oli useita tilastollisesti merkitseviä termejä. Selitysasteet jäivät hyvin pieniksi. Mallin selittävät muuttujat eivät selitä hyvin tutkittavaa ilmiötä, koska ne eivät korreloi tarpeeksi paljoa vastemuuttujan kanssa. Mallia käytetään mallintamaan selitettävää osaa puvuista.

Taulukossa 12 on neljä saraketta. Taulukon sarakkeiden sisältö on kuvattuna alapuolelle olevassa listassa:

- *Erät:* Erien lukumäärä selitettävässä osassa.
- *Yli 0,5:* Niiden erien lukumäärä, jotka kuuluivat yli 50 % todennäköisyydellä mallinnettuun logistiikkaketjuun.

- *Kokonaisarvo miljoonaa euroa:* Selitettävän osan euromääräinen arvo vaateryhmittäin.
- *Euroarvo korjattu miljoonaa euroa:* Niiden erien kokonaisarvo, jotka saivat yli 50 % todennäköisyyden kuulua mallinnettuun logistiikkaketjuun.

Taulukko 12. Pukujen tuonnin keskeisiä tietoja

erät	erät yli 0,5	kokonaisarvo miljoonaa	euroarvo korjattu miljoonaa euroa
5456	4	21.6	1.4

Ryhmässä puvut neljä erää sai yli 50 % todennäköisyyden kuulua mallinnettuun logistiikkaketjuun. Näiden arvo oli 1,4 miljoonaa euroa. Koko vaateryhmän arvosta 1,4 miljoonaa euroa on noin 6,5 %.

Alusvaatteiden tuontia mallinnettiin logistisella regressiomallilla. Aineistosta rajattiin alkuperämaiksi Aasia, lähtömaaksi Ruotsi ja vaateryhmäksi alusvaatteet.

Logistisen regressiomallin tilastollisesti merkitsevät muuttujat ja niiden interaktiot on kuvattu alla. Mallin parametriestimaatit ja tilastollisia tunnuslukuja on esitetty taulukossa 13.

$$\text{logit } P(Y) = \text{vakio} + \text{yarvo} + \text{sesonki} + \text{yarvo} * \text{sesonki} + \text{paino}$$

Taulukko 13. Mallin parametriestimaatit, kun alkuperämaana Aasia ja ryhmänä alusvaatteet

	Analysis	of Maximum	Likelihood	Estimates	
parametri	DF	estimaatti	SE	Chi-Square	p - ChiSq
Intercept	1	-2.0748	0.0164	15972.4058	<.0001
yarvo	1	0.0115	0.000554	430.3102	<.0001
sesonki	1	0.0529	0.0231	5.254	0.0219
yarvo*sesonki	1	0.00215	0.000776	7.651	0.0057
paino	1	0.000073	0.000022	11.1183	0.0009

Taulukko 14. Mallin selitysasteet vaateryhmässä alusvaatteet

R-Square	0.0108	Max-rescaled	R-Square	0.0194
----------	--------	--------------	----------	--------

Malliin löytyi useita tilastollisesti merkitseviä termejä. Selitysasteet jäivät kuitenkin hyvin pieniksi. Mallin selittävät muuttujat eivät selitä hyvin tutkittavaa ilmiötä. Mallia käytetään mallintamaan selitettävää osaa alusvaatteista.

Taulukko 15. Alusvaatteiden tuonnin keskeisiä tietoja

erät	yli 0,5	kokonaisarvo miljoonaa euroa	euroarvo korjattu miljoonaa euroa
4825	449	18.5	1.1

Alusvaatteiden eristä 449 sai yli 50 % todennäköisyyden kuulua mallinnettuun logistiikkaketjuun. Näiden erien arvo oli 1,1 miljoonaa euroa. Koko vaateryhmän arvosta 1,1 miljoonaa on noin 5,9 %.

Mallinnettiin muiden tekstiilien tuontia logistisella regressiomallilla. Aineistosta rajattiin alkuperämaiksi Aasia, lähtömaaksi Ruotsi ja vaateryhmäksi muut tekstiilit.

Logistisen regressiomallin tilastollisesti merkitsevät muuttujat ja niiden interaktiot on kuvattu alla. Mallin parametriestimaatit ja tilastollisia tunnuslukuja on esitetty taulukossa 16.

$$\text{logit } P(Y) = \text{vakio} + \text{yarvo} + \text{kulj} + \text{yarvo} * \text{kulj} + \text{sesonki} + \text{yarvo} * \text{sesonki} + \text{paino}$$

Taulukko 16. Mallin parametriestimaatit, kun alkuperämaana Aasia ja ryhmänä muut tekstiilit

	Analysis	of Maximum	Likelihood	Estimates	
parametri	DF	estimaatti	SE	Chi-Square	p - ChiSq
Intercept	1	-2.9899	0.1558	368.0710	<.0001
yarvo	1	0.00405	0.000742	29.8876	<.0001
kulj	1	2.845	0.1590	320.2117	<.0001
yarvo*kulj	1	-0.1501	0.00259	3358.6944	<.0001
sesonki	1	0.1693	0.0334	25.7739	<.0001
yarvo*sesonki	1	-0.0109	0.00340	10.230	0.0014
paino	1	0.00269	0.000179	225.0146	<.0001

Taulukko 17. Mallin selitysasteet vaateryhmässä muut tekstiilit

R-Square	0.1309	Max-rescaled	R-Square	0.2228
----------	--------	--------------	----------	--------

Malliin löytyi useita tilastollisesti merkitseviä termejä. Selitysasteet nousivat varsin suuriksi, kun vertaa muihin vaateryhmiin. Mallin selittävät muuttujat selittävät melko hyvin tutkittavaa ilmiötä. Mallia käytetään mallintamaan selitettävää osaa muista tekstiileistä.

Taulukko 18. Muiden tekstiilien tuonnin keskeisiä tietoja

erät	yli 0,5	kokonaisarvo miljoonaa euroa	euroarvo korjattu miljoonaa euroa
3685	497	16.4	11.2

Muiden tekstiilien eristä 497 erää sai yli 50 % todennäköisyyden kuulua

mallinnettuun logistiikkaketjuun. Näiden arvo oli 11,2 miljoonaa euroa. Koko vaateryhmän arvosta 11,2 miljoonaa on noin 68 %.

6.3 Muiden vaateryhmien tuloksia

Tässä luvussa esitetään yleisellä tasolla muut vaateryhmät, joita ei vielä käsitelty. Taulukossa 19 on kuvattuna näiden ryhmien euromääräiset alkuperämaajakaumat.

Taulukko 19. Muiden ryhmien tuonnin keskeisiä tietoja

Tavararyhmä	erät	yli 0,5	kokonaisarvo miljoonaa	euroarvo korjattu
Takit	1585	0	0	0
Paidat	3460	27	10	0.4
Vauvanvaatteet	198	0	0.7	0
Urheiluvaatteet	1152	1	30.7	0.05
Muut vaatteet	2647	12	4.9	0.02

Edellä kuvattujen ryhmien alkuperäiset euromääräiset jakaumat muuttuivat hyvin vähän.

6.4 Korjattu alkuperämaajakauma

Taulukossa 20 on kuvattuna lähtötilanteen alkuperämaajakauma: alkuperämaiden euromäärät, erät ja prosenttiosuudet. Maahan AA on niputettu kaikki Aasian maat.

Taulukko 20. Alkuperäinen alkuperämaajakauma

maa	summa euroa	erät	prosenttia / kokonaisarvosta
AA	96 257 756	106 783	55,00 %
SE	79 540 487	23 008	45,00 %

Taulukossa 21 on kuvattuna korjattu alkuperämaajakauma: alkuperämaiden euromäärät, erät ja prosenttiosuudet. Maahan AA on niputettu kaikki Aasian maat.

Taulukko 21. Korjattu alkuperämaajakauma

maa	summa euroa	erät	prosenttia / kokonaisarvosta
AA	110 427 756	107 773	62,80 %
SE	65 370 487	22 018	37,20 %

Korjatussa alkuperämaajakaumassa 990 erän osalta alkuperämaatieto muuttui. Erien arvo oli 14,2 miljoonaa euroa ja tämä oli 7,8 % vaatetuonnin kokonaisarvosta.

7 JOHTOPÄÄTÖKSET JA POHDINTA

Tutkielmassa tutkittiin Suomen vaatetuonnin alkuperämaajakaumaa vuodelta 2005. Ajallisesti eniten aikaa kului aineiston muokkaamiseen oikeanlaiseksi tilastollisia menetelmiä varten. Kyseisenlaista tutkimusta ei ole tehty aikaisemmin. Aineiston muokkaamiseen ja tilastollisiin ajoihin käytettiin tutkielmassa SAS 8.2 ohjelmaa.

Tutkielman aineisto jaettiin kahteen osaan. Nämä aineiston osat olivat selittävä osa, jonka alkuperämaatieto oletettiin oikein kirjatuksi sekä selitettävä osa, jonka alkuperämaatieto oletettiin väärin kirjatuksi. Selittäväälle osalle sovitettiin logistinen regressiomalli, jonka parametriestimaatteja käytettiin selitettävälle osalle. Näin saatiin jokaiselle vaate-erälle todennäköisyysarvo, että alkuperämaatieto olisi Aasia Ruotsin sijaan.

Tuloksissa tarkasteltiin, kuinka saatu tilastollinen malli jakoi selitettävän osan aineiston uudelle alkuperämaajakaumalle. Lähtötilanteen selitettävän alkuperämaajakauman arvosta 79,5 miljoonaa euroa sai noin 14,2 miljoonaa euroa uuden alkuperämaatiedon.

Eniten alkuperämaatieto muuttui vaate-ryhmässä muut tekstiilit. Jakauma muuttui yli 68 % alkuperäisestä arvosta, ja ainakin tämän ryhmän osalta on syytä olettaa, että alkuperämaatieto olisi osittain väärin kirjattu. Muissa vaate-ryhmissä muutos oli hyvin vähäistä ja niissä myös selitysasteet olivat varsin pieniä.

Alkuperämaa ongelmaa olisi huomattavasti helpompi tutkia, jos olisi

tiedossa, missä vaate-erissä alkuperämaatietao on korvattu lähtömaatiedolla. Näin ollen ongelma palautuisi imputointiongelmaksi.

Tuloksista käy ilmi, että alkuperämaatietao olisi syytä epäillä laadullisesti heikoksi. Tästä syystä sen keräämismenetelmiä tulisi kehittää, jos tietoa on tarkoitus tilastoida myös tulevaisuudessa.

LIITTEET

Liite 1

Aineiston määrittäjiä:

Tavaraerän numero: Viranomaisen täyttää. Kohtaan merkitään tilastoilmoituksen järjestysnumero asiakirjan arkistointia varten. (Intrastat, 2005)

Kauppatapahtuman luonne: Ilmoitetaan kauppatapahtuman luonne. Esim. Normaali osto tai myynti = koodi 11. (Intrastat, 2005)

Jakso: Merkitään ajanjakso, jota tilastoilmoitus koskee. (Intrastat, 2005)

Toinen paljous ja sen laji: Ilmoitetaan sovelletun cn-tavaranimikkeen edellyttämä toinen paljous, esim. litra tai kappalemäärä. (Intrastat, 2005)

Laskutusarvo: Laskutusarvo ilmoitetaan kokonaislukuna, ilman desimaaleja (täysinä euroina ei sentteinä). (Intrastat, 2005)

Kuljetusmuoto: Kuljetusmuoto määräytyy sen mukaan, millä aktiivisella kulkuneuvolla tavara tuodaan tai viedään Suomen rajan yli. Mikäli kuljetusmuoto ei ole tiedossa ilmoitetaan todennäköisin kuljetusmuoto. (Intrastat, 2005)

Nettopaino: Ilmoitetaan nimike-erän puhdas nettopaino kilogrammoina (kg). Nettopaino tarkoittaa tavaran painoa ilman päällystä tai pakkausta. (Intrastat, 2005)

Laskutusarvo: Tuonnissa on yhteisöhankinnan verotusarvo, jolla tarkoitetaan myyjän ja ostajan väliseen sopimukseen perustuvaa hintaa, joka sisältää kaikki myyjän ja ostajalta perimät hinnannlisät. (Intrastat, 2005)

Tavaranimike: Merkitään yhdistetyn nimikkeistön (CN-luokitus) mukainen tavaranimike. CN-luokitus kattaa Suomen käyttötariffinimikkeistön (TARIC) 8 ensimmäistä numeroa. (Intrastat, 2005)

Alkuperämaat tutkielmassa (Aasia):

Bangladesh (BD), Indonesia (ID), Intia (IN), Kiina (CN) ja Pakistan (PK).

Alla vaateryhmien CN4 luokat:

Ryhmä 1:

Takit ryhmään kuuluu CN4 -tasolla luokat 6101, 6202, 6201, ja 6102.

Ryhmä 2:

Puvut ryhmään kuuluu CN4 -tasolla luokat 6103, 6203, 6104 ja 6204.

Ryhmä 3:

Paidat ryhmään kuuluu CN4 -tasolla luokat 6105, 6106, 6110, 6205 ja 6206.

Ryhmä 4:

Vauvanvaatteet ryhmään kuuluu CN4 -tasolla luokat 6111 ja 6209.

Ryhmä 5:

Alusvaatteet ryhmään kuuluu CN4 -tasolla luokat 6107, 6108, 6109, 6115, 6212, 6207 ja 6208.

Ryhmä 6:

Urheiluvaatteet ryhmään kuuluu CN4 -tasolla luokat 6112 ja 6211.

Ryhmä 7:



Muut vaatteet ryhmään kuuluu CN4 -tasolla luokat 6113, 6114, 6116, 6117, 6210, 6213, 6214, 6215, 6216 ja 6217.

Ryhmä 8:

Muut tekstiilit ryhmään kuuluu CN4 -tasolla luokat väliltä 6300 - 6399.

Liite 2

Tulli ilmoitus:

		LOMAKE II Tuonti [1]																								
Tiedonantovelvollinen FI 6663662-2 Import-Export Finland		Tilastojakso 2007-01		Ilmoitusnumero ja tila 07-11-147-EL0-005 JUS																						
Asiamies FI 7654321-2 Itella Demoyritys		Viite ohje <input type="text"/>																								
Ilmoitus, jossa ei ole yhtään nimike-erää, tulkitaan nolailmoitukseksi.																										
Tavaranimike <input type="text"/> <input type="button" value="Eae"/>		Tavarankuvaus (pöytä) <input type="text"/>																								
Lähetysmaa <input type="text" value=" <Valitse >"/>																										
Alkuperämaa ohje <input type="text" value=" <Valitse >"/>																										
Kauppatahtumuksen luonne <input type="text" value=" 11 - Suora osto,myynti"/>																										
Kuljetusmuoto <input type="text" value=" 1 - Merikuljetus (ml. auro- ja junaleutta-kuljetus)"/>																										
Nettopaino kg ohje <input type="text"/> <input type="text" value=" pako linter"/>		Toinen paljous ohje <input type="text"/>																								
Laskutusarvo ohje <input type="text"/> EUR		Tilastoarvo ohje <input type="text"/> EUR																								
Laskutusarvo muussa valuutassa ohje <input type="text"/>		Tilastoarvo muussa valuutassa ohje <input type="text"/>																								
Syötä askutus- ja tilastoarvot euroissa tai muussa valuutassa. Jos syötät arvon muussa valuutassa kuin euroissa, lasketaan euroarvo automaattisesti.																										
<input type="button" value="Lisää listaan"/> <input type="button" value="Poista listasta"/> <input type="button" value="Muuta valittua"/> <input type="button" value="Tyhjennä kertät"/>																										
<table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th>#</th> <th>Nimike</th> <th>Lä»</th> <th>Al»</th> <th>Kt</th> <th>Km</th> <th>Nettop.</th> <th>2. palj.</th> <th>yks.</th> <th>Laskutusarvo</th> <th>Tilastoarvo</th> </tr> </thead> <tbody> <tr> <td>0</td> <td>--</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> </tbody> </table>					#	Nimike	Lä»	Al»	Kt	Km	Nettop.	2. palj.	yks.	Laskutusarvo	Tilastoarvo	0	--									
#	Nimike	Lä»	Al»	Kt	Km	Nettop.	2. palj.	yks.	Laskutusarvo	Tilastoarvo																
0	--																									
Erien laskutusarvot yhteensä: <input type="text"/> EUR																										
<input type="button" value="Hyväksy ja lähetä"/> <input type="button" value="Tallenna lähettämättä"/> <input type="button" value="Palaa pääsivulle"/>																										

LÄHTEET

Cox, D. R. & Snell, E. J. (1989) Analysis of binary data. Second edition. London: Chapman and Hall.

Hosmer, D. W. Jr & Lemeshow, S. (1989). Applied logistic regression. New York: Wiley.

Hosmer, D. W. Jr & Lemeshow, S. (2000). Applied logistic regression. New York: Wiley.

Intrastat Suomessa vuonna 2006. (2005). Forssan kirjapaino Oy. Forssa.

Lehtonen, R. & Pahkinen, E. (2004). Practical methods for design and analysis of complex Surveys. Second edition. Chichester: Wiley.

Liang, K. Y. & Zeger, S. L. 1986. Longitudinal data analysis using generalized linear models. *Biometrika*, Vol 73, no. 1. s. 13-22.

Nagelkerke, N. J. D. 1991. Miscellanea: A note on a general definition of the coefficient of determination. *Biometrika*, Vol 78, no. 3. s. 691-692.

Ortega, J. M. & Rheinboldt, W. C. (2000). Iterative solutions of nonlinear equations in several variables. Philadelphia, PA: SIAM.

Suomen Pankki. 2005. Lehdistötiedote nro 5 Maksutasetilastointi monipuolistuu. Suomen Pankki [online][Viitattu 04.12.2007]
<http://www.suomenpankki.fi/fi/suomen_pankki/ajankohtaista/tiedotteet/2005/tiedote5_2005.htm>

Tulli 2006a. Ulkomaankauppatilaston käyttötarkoitus ja keskeiset käsitteet.

Tulli [online]. [Viitattu 24.10.2006] Saatavilla [www-osoitteessa](http://www.tulli.fi)

http://www.tulli.fi/fi/05_Ulkomaankauppatilastot/01_SVT/02_Kasikirja/laatu2004.jsp

Tulli. 2006b. Tavaroiden ulkomaankaupan kuukausijulkaisu. Tulli [online].

[Viitattu 26.11.2006] Saatavilla [www-osoitteessa](http://www.tulli.fi)

<http://www.tulli.fi/fi/05_Ulkomaankauppatilastot/01_SVT/03_Laatuselosteet/kk-julkaisu.jsp>

Tulli. 2007. CN-Nimikkeistö. Tulli [online]. [Viitattu 13.10.2007] Saatavilla

[www-osoitteesta](http://www.tulli.fi)

<http://www.tulli.fi/fi/05_Ulkomaankauppatilastot/03_CN_nimikkeisto/index.jsp>