



Open your mind. LUT.
Lappeenranta University of Technology

LAPPEENRANNAN TEKNILLINEN YLIOPISTO

Kauppätieteellinen tiedekunta

Kandidaatintutkielma

Talousjohtaminen

Endogeenisuus lineaarisessa regressiossa

Endogeneity in linear regression

Elsa Nyman

17.12.2012

Sisällysluettelo

1	Johdanto.....	1
2	Lineaarinen regressiomalli.....	2
3	Pienimmän neliösumman menetelmä	5
4	Endogeenisuusongelma	7
4.1	Tilanteita joissa selittävät muuttuja ja virhetermi ovat korreloituneet	7
4.1.1	Puuttuvat selittävät muuttujat.....	7
4.1.2	Samanaikaisuusharha	8
4.1.3	Mittausvirheet.....	9
4.2	Miksi pienimmän neliösumman menetelmää ei voida käyttää?	11
4.3	Momenttimenetelmä	12
4.3.1	Instrumenttimuuttujamenetelmä eli kaksivaiheinen pienimmän neliösumman menetelmä	14
4.3.2	Instrumenttimuuttujan pätevyyden testaaminen.....	16
4.4	Endogeenisuuden testaaminen	17
5	Esimerkki endogeenisuusongelmasta: tutkimus siirtomaiden instituutioiden vaikutuksesta maiden taloudelliseen suorituskyykyyn	19
5.1	Tutkimuksen taustaa	19
5.2	Tutkimuksen hypoteesit.....	19
5.3	Tutkimuksen aineisto ja muuttujat	21
5.4	Tutkimuksessa ilmenevät ongelmat.....	22
5.4.1	Mallin lineaarisuus	22
5.4.2	Puuttuvat selittävät muuttujat.....	23
5.4.3	Samanaikaisuusharha	23
5.4.4	Mittausvirhe.....	23
5.5	Endogeenisuuden korjaus	24
5.6	Endogeenisuuden testaaminen	25
6	Tutkimustulokset	25

7	Johtopäätökset	29
	LIITTEET	32
	Liite 1: Tutkimuksen aineisto ja muuttujat.....	32

1 Johdanto

Tämän kandidaatintutkielman tarkoituksena on tarkastella lineaarisessa regressiomallissa esiintyvää endogeenisuusongelmaa. Tutkielmassa lähdetään liikkeelle lineaarisen regressiomallin ja sen parametrien yleisimmän estimointimenetelmän, pienimmän neliösumman menetelmän, määrittelystä. Jos jokin lineaarisen regressiomallin perusoletuksista ei ole voimassa, pienimmän neliösumman menetelmällä ei saada luotettavia estimaatteja mallin parametreista (Hill, Griffiths & Lim, 2012). Tässä tutkielmassa keskitytään lineaarisen regressiomallin oletukseen, jonka mukaan mallin selittävän muuttujan ja virhetermin välillä ei saa olla korrelaatiota. Jos kyseiset muuttujat ovat korreloituneet keskenään, mallissa on endogeenisuusongelma ja parametrien estimointiin ei voida käyttää pienimmän neliösumman menetelmää (Hill et al. 2012). Pienimmän neliösumman menetelmä esitetään lyhyesti kappaleessa 3.

Kappaleessa 4 kuvataan yleisimmät tilanteet, jotka aiheuttavat regressiomallin endogeenisuusongelman: puuttuvat selittävät muuttujat, samanaikaisuusharha ja mittausvirhe (Hill et al. 2012). Berkowitz & Hoekstra (2010) ovat tutkineet Yhdysvaltain parhaiden yksityisten lukioiden opiskelijoiden todennäköisyyttä päästä parhaimpiin yliopistoihin. Kyseisessä tutkimuksessa esitetään, että saadut opiskelupaikat riippuvat kuitenkin myös muista seikoista, joita ei ole otettu mallissa huomioon. Jos jokin näistä seikoista on korreloitunut tarkasteltavan henkilön lukiokoulutuksen tasoa mittaavan muuttujan kanssa, mallissa on endogeenisuusongelma. Tätä puuttuvien selittävien muuttujien ongelmaa eli ”omitted variables” – ongelmaa tarkastellaan kappaleessa 4.1.1.

Kappaleessa 4.1.2. esitetään moniyhtälömalleissa esiintyvä samanaikaisuusharha eli ”simultaneous bias” – ongelma hyödykkeen kysyntä- ja tarjontayhtälöiden avulla. Hyödykkeen hinnan ja vaihdetun määrän välinen kausaalisuus on epäselvä: ei voida selkeästi sanoa vaikuttaako hyödykkeen hinta vaihdettuun määrään vai päinvastoin.

Taloudellisia muuttujia on vaikea mitata. Tällöin muuttuja on voitu korvata toisella muuttujalla, joka ei mittaa täsmällisesti haluttua asiaa. Tämä johtaa siihen, että mallissa on mittausvirhe eli ”errors-in-variables” – ongelma. (Hill et al. 2010) Kappalees-

sa 4.1.3. tarkastellaan esimerkkiä, jossa henkilön tulotason ja koulutustason välillä on endogeenisuusongelma johtuen koulutustason vaikeasta mitattavuudesta (Greene, 2003).

Edellisten esimerkkien perusteella voidaan todeta, että endogeenisuusongelmaa esiintyy hyvinkin arkipäiväisissä tilanteissa. Näiden tilanteiden tunteminen on erityisen tärkeää kvantitatiivisessa tutkimuksessa, jotta saadaan oikeanlaisia tuloksia. Endogeenisuusongelman tapauksessa pienimmän neliösumman menetelmällä saadaan harhaisia estimaatteja mallin parametreista. Mikäli endogeenisuusongelma havaitaan, voidaan parametrit estimoida vaihtoehtoisella estimointimenetelmällä, esimerkiksi kappaleessa 4.3.1. esitetyllä instrumenttimuuttujamenetelmällä. (Hill et al. 2012)

Acemoglu, Johson ja Robinson (2000) ovat tutkineet siirtomaihin perustettujen instituutioiden ja maiden taloudellisen suorituskyvyn välisessä suhteessa ilmenevää endogeenisuusongelmaa. Tutkimus koostuu 71 maasta, jotka ovat olleet eurooppalaisien siirtomaita 1900-luvulla. Acemoglu et al. (2000) esittävät, että instituutioita mitaava muuttuja on mitattu virheellisesti ja siten korreloitunut mallin virhetermin kanssa. Tällöin pienimmän neliösumman menetelmällä saadut tulokset ovat harhaisia. Kappaleessa 5 analysoin kyseistä tutkimusta kappaleen 4 teoriarungon mukaisesti. Tutkimuksen mukaisten testien suorittamiseen olen käyttänyt tilastollista ohjelmistoa.

Kappaleeseen 6 olen koonnut suorittamieni testien tulokset, joista nähdään, että mallissa ilmenee endogeenisuusongelma. Ilman tietoa endogeenisuutta aiheuttavista tekijöistä, tutkimuksen tuloksista olisi todennäköisesti tehty virheellisiä päätelmiä. Kappaleessa 7 esitetään tutkielman johtopäätökset ja jatkotutkimusaiheet.

2 Lineaarinen regressiomalli

Taloustieteissä ollaan kiinnostuneita yhden muuttujan vaihtelun vaikutuksesta toiseen muuttajaan. Yleinen tutkimuksen kohde on hinnan muutoksen vaikutus hyödykkeen vaihdettuun määrään. Lisäksi taloustieteissä pyritään ennustamaan yhden

muuttujan arvo, kun toisen muuttujan arvo tunnetaan. Tällaisia ongelmia voidaan tutkia regressiomallin avulla. (Hill et al. 2012)

Regressiomallissa yhtälön vasemmalla puolella on selitettävä muuttuja ja oikealla puolella siihen vaikuttavat tekijät. Yhden selittävän muuttujan lineaarinen regressiomalli on muotoa:

$$y = \beta_0 + \beta_1 x + e,$$

missä y = selitettävä muuttuja, β_0 = vakiotermi, β_1 = kulmakerroin, x = selittävä muuttuja ja e = virhetermi. Vakiotermi kertoo sen pisteen, jossa regressio leikkaa koordinaatiston vertikaalisen akselin. Kulmakerroin kuvaa selittävän muuttujan vaihtelun vaikutusta selitettävään muuttujaan. Virhetermi sisältää muut selitettävään muuttujaan vaikuttavat tekijät. (Hill et al. 2012)

Usean selittävän muuttujan lineaarinen regressiomalli on muotoa:

$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + e,$$

missä y = selitettävä muuttuja, β_0 = vakiotermi, (x_1, \dots, x_k) = selittävät muuttujat ja e = virhetermi. Parametrit $(\beta_1, \dots, \beta_k)$ kuvaavat kunkin selittävän muuttujan vaikutusta selitettävään muuttujaan. (Roberts & Whited, 2011)

Oletetaan esimerkiksi, että henkilön korkeampi koulutustaso mahdollistaa korkeamman tulotason. Lineaarinen regressiomalli muodostetaan seuraavasti:

$$\text{tulotaso} = \beta_0 + \beta_1 \text{koulutus} + e,$$

missä β_0 = vakiotermi, β_1 = kulmakerroin ja e = virhetermi. Vakiotermi kertoo tulotason alhaisimman mahdollisimman arvon. Kulmakerroin kuvaa sitä, kuinka paljon koulutustason muutos vaikuttaa tulotasoon. (Greene, 2003)

Selitettävään muuttujaan vaikuttavat tekijät jaetaan kahteen komponenttiin: systemaattiseen komponenttiin $\beta_0 + \beta_1 x$ ja satunnaiseen komponenttiin eli virhetermiin e .

Systemaattinen komponentti on selitettävän muuttujan odotusarvo eli selitettävän muuttujan saamien arvojen keskiarvo, joka on painotettu arvojen todennäköisyyksillä. Systemaattisen komponentin odotusarvo on ehdollinen, sillä se riippuu selittävän muuttujan saamista arvoista. Ehdollinen odotusarvo esitetään seuraavasti:

$$E(y|x) = \beta_0 + \beta_1 x,$$

missä $E(y|x)$ = selitettävän muuttujan odotusarvo selittävän muuttujan suhteen, β_0 = vakiotermi, β_1 = kulmakerroin ja x = selittävä muuttuja. Systemaattinen komponentti on matemaattinen odotus, minkä vuoksi se ei ole satunnainen. (Hill et al. 2012)

Selitettävä muuttuja ja satunnainen komponentti sen sijaan ovat satunnaisia. Tämä tarkoittaa, että kyseisten muuttujien arvoja ei tiedetä ennen tutkimuksen suorittamista. Satunnainen komponentti esitetään selitettävän muuttujan ja systemaattisen komponentin erotuksena:

$$e = y - E(y|x) = y - \beta_0 - \beta_1 x$$

Yhtälöstä nähdään, että selitettävä muuttuja ja satunnainen komponentti eroavat toisistaan vain ei-satunnaisen systemaattisen komponentin verran. Tällöin satunnaisen komponentin odotusarvo on myös riippuvainen selittävän muuttujan saamista arvoista. Satunnaisen komponentin ehdollinen odotusarvo esitetään seuraavasti:

$$E(e|x) = E(y|x) - (\beta_0 + \beta_1 x) = 0,$$

missä $E(e|x)$ = virhetermin odotusarvo selittävän muuttujan suhteen, $E(y|x)$ = selitettävän muuttujan odotusarvo selittävän muuttujan suhteen ja $(\beta_0 + \beta_1 x)$ = systemaattinen komponentti. Yhtälöstä nähdään, että virhetermin odotetaan olevan nolla. Tällöin lineaarinen regressio on muotoa $y = \beta_0 + \beta_1 x$ ja voidaan olettaa, että selitettävään muuttujaan vaikuttaa vain selittävä muuttuja x . Koska virhetermi kuvaa kaikkia niitä muita selitettävään muuttujaan vaikuttavia tekijöitä, virhetermin poikkeaminen nolasta voi viitata siihen, että mallista puuttuu merkittäviä selittäviä muuttujia. Toisaalta, jos selittävä muuttuja on mitattu virheellisesti, todellisen selittävän muuttujan vaikutus selitet-

tävään muuttujaan näkyy virhetermissä. Tällaisissa tapauksissa selittävää muuttujaa sanotaan endogeeniseksi. (Hill et al. 2012)

3 Pienimmän neliösumman menetelmä

Yhden muuttujan lineaarisessa regressiomallissa tarkoituksena on ratkaista parametrit β_0 ja β_1 eli mallin vakiotermin ja kulmakerroin. Pienimmän neliösumman menetelmässä estimoitu suora muodostuu siten, että vertikaalisen etäisyyden neliö jokaista aineiston arvoa vastaavasta pisteestä suoralle on mahdollisimman pieni. Vertikaalisten etäisyyksien neliöön korottamisella vältetään suurten negatiivisten etäisyyksien vaikutukselta. Saadun suoran vakiotermin b_0 ja kulmakerroin b_1 ovat pienimmän neliösumman estimaatit parametreista β_0 ja β_1 . Estimoitu suora on muotoa:

$$\hat{y}_i = b_0 + b_1 x_i,$$

missä \hat{y}_i = estimoitu selitettävä muuttuja, b_0 = vakiotermin estimaatti, b_1 = kulmakerroin estimaatti ja x_i = selittävä muuttuja. Todellisen selitettävän muuttujan ja estimoidun selitettävän muuttujan etäisyyttä sanotaan pienimmän neliösumman residuaaliksi ja se esitetään seuraavan yhtälön avulla:

$$\hat{e}_i = y_i - \hat{y}_i = y_i - b_0 - b_1 x_i,$$

missä \hat{e}_i = pienimmän neliösumman residuaali, y_i = selitettävä muuttuja ja \hat{y}_i = estimoitu selitettävä muuttuja. Pienimmän neliösumman estimaateille b_0 ja b_1 on ominaista, että niiden residuaalien neliösumma on pienempi kuin minkä tahansa muun suoran residuaalien neliösumma. (Hill et al. 2012)

Pienimmän neliösumman estimaatit saadaan etsimällä kaikista mahdollisista parametrien β_0 ja β_1 arvoista piste (b_0, b_1) , jossa neliösummafunktio $S(\beta_0, \beta_1)$ minimoituu seuraavan kaavan mukaisesti:

$$S(\beta_0, \beta_1) = \sum_{i=1}^N (y_i - \beta_0 - \beta_1 x_i)^2$$

Kyseisessä pisteessä saadaan estimaattorit b_0 ja b_1 :

$$b_0 = \bar{y} - b_1 \bar{x}$$

$$b_1 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2},$$

missä $\bar{y} = \sum y_i / N$ ja $\bar{x} = \sum x_i / N$. Ne ovat selitettävän muuttujan y ja selittävän muuttujan x havaintojen otoskeskiarvoja. Vakiotermin b_0 saadaan selitettävän muuttujan otoskeskiarvoa selittävästä yhtälöstä. Kulmakerrointa b_1 estimoivan yhtälön osoittajassa lasketaan yhteen tulo, joka koostuu kahdesta komponentista: selittävän muuttujan x havaintojen poikkeamasta otoskeskiarvosta \bar{x} ja selitettävän muuttujan y havaintojen poikkeamasta otoskeskiarvosta \bar{y} . Näiden kahden komponentin tuloa sanotaan kovarianssiksi. Yhtälön nimittäjä koostuu osoittajan ensimmäisen komponentin neliöstä eli selittävän muuttujan havaintojen varianssien summasta. Pienimmän neliösumman estimaatit parametreille β_0 ja β_1 saadaan, kun selitettävän muuttujan otosarvot y_i ja selittävän muuttujan otosarvot x_i sijoitetaan yllä oleviin yhtälöihin. (Hill et al. 2012)

Lineaarisen regressiomallin oletusten pitäessä paikkansa, pienimmän neliösumman estimaattori on paras lineaarinen harhaton estimaattori. Jos jokin oletus ei pidä paikkansa, sanotaan että pienimmän neliösumman estimointimenetelmä ei ole konsistentti eli ristiriidaton. Endogeenisuusongelman tapauksessa satunnainen selittävä muuttuja korreloi virhetermin kanssa, jolloin pienimmän neliösumman menetelmä ei ole paras lineaarinen harhaton estimointimenetelmä. Tällöin tulisi käyttää jotakin toista estimointimenetelmää. (Hill et al. 2012)

4 Endogeenisuusongelma

Lineaarisessa regressiossa oletuksena on, että virhetermin odotusarvo on nolla. Jos oletus pitää paikkansa, selittävä muuttuja ei korreloi virhetermin kanssa ja sen sanotaan olevan eksogeeninen. Jos virhetermin odotusarvo poikkeaa nolasta ja se korreloi selittävän muuttujan kanssa, sanotaan selittävää muuttujaa endogeeniseksi. Jos mallissa on endogeenisuusongelma, pienimmän neliösumman estimaattorit eivät ole konsistentteja. (Hill et al. 2012)

4.1 Tilanteita joissa selittävä muuttuja ja virhetermi ovat korreloituneet

4.1.1 Puuttuvat selittävät muuttujat

Puuttuvilla selittäville muuttujilla tarkoitetaan sellaisia muuttujia, joilla on merkittävä vaikutus selitettävään muuttujaan, mutta jotka on jätetty mallin ulkopuolelle. Tällöin kyseiset muuttujat eivät näy selittämissä muuttujissa, vaan virhetermissä. Jos puuttuvat selittävät muuttujat eivät korreloi mukaan otettujen selittävien muuttujien kanssa, pienimmän neliösumman estimaattorit ovat konsistentteja ja harhattomia. Endogeenisuusongelma syntyy, jos kyseiset muuttujat ovat korreloituneita keskenään. Ongelmaa voidaan kuvata seuraavan yhtälön mukaisesti:

$$y = \beta_0 + \beta_1 x + kw + e,$$

missä y = selitettävä muuttuja, β_0 = vakiotermi, β_1 = kulmakerroin, x = selittävä muuttuja, w = havaitsematon selittävä muuttuja, k = havaitsemattoman selittävän muuttujan kerroin ja e = virhetermi. Malli voidaan muotoilla seuraavasti:

$$y = \beta_0 + \beta_1 x + v,$$

missä $v = kw + e$ eli yhdistetty virhetermi. Jos havaitsematon selittävä muuttuja on korreloitunut mallin selittävän muuttujan kanssa, myös yhdistetty virhetermi on korreloitunut selittävän muuttujan kanssa. Tällöin voidaan olettaa, että korreloitunut selit-

tävä muuttuja on endogeeninen ja pienimmän neliösumman estimointimenetelmä tuottaa epäkonsistenttejä estimaattoreita kaikista mallin parametreista. (Roberts & Whited, 2011)

Berkowitz & Hoekstra (2010) ovat tutkineet, kuinka parhaat yksityiset lukiot vaikuttavat yliopistopaikkoihin Yhdysvalloissa. Tutkimuksen tulokset osoittavat, että yksityisten koulujen vaikutus parhaiden yliopistopaikkojen saamiseen on suurempi kuin julkisten koulujen. Tätä yhteyttä voidaan kuvata seuraavan yhtälön avulla:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + e,$$

missä x_1 = keskikoulun keskiarvo, x_2 = muuttuja, joka mittaa sitä, onko tarkasteltava henkilö käynyt yksityistä koulua, x_3 = pääsykoepisteet, x_4 = muuttuja, joka kertoo onko joku tarkasteltavan henkilön sukulaisista käynyt yksityistä koulua, x_5 = vektori, joka kuvaa muita selitettävään muuttujaan vaikuttavia muuttujia (esimerkiksi naapurusto, tulot, hakuvuosi), $(\beta_0, \dots, \beta_5)$ = mallin parametrit ja e = virhetermi. Mallin endogeenisuusongelma syntyy, jos jokin muuttujista (x_2, \dots, x_5) on korreloitunut muuttujan x_1 kanssa. Berkowitzin ja Hoekstran (2010) tutkimuksen tuloksista voidaan todeta, että yksityisten lukioiden oppilaat pääsevät todennäköisemmin parempiin yliopistoihin kuin julkisten lukioiden. Tuloksista havaitaan myös, että mikään havaitsemattomista muuttujista ei korreloi tarkasteltavan henkilön koulutustasoa mittaavan muuttujan kanssa. Tällöin mallissa ei ole puuttuvien selittävien muuttujien aiheuttamaa endogeenisuusongelmaa ja pienimmän neliösumman menetelmä on konsistentti.

4.1.2 Samanaikaisuusharha

Samanaikaisuusharha esiintyy moniyhtälömalleissa, kun selitettävä muuttuja y ja jokin selittävästä muuttujista x_k määräytyvät tasapainossa siten, että voidaan sanoa sekä muuttujan x_k vaikuttavan muuttujaan y että muuttujan y vaikuttavan muuttujaan x_k (Roberts & Whited, 2011). Samanaikaisuusharhaa voidaan havainnollistaa markkinoiden kysynnän ja tarjonnan tasapainon avulla.

$$\text{Kysyntä: } Q = \alpha_1 P + \alpha_2 I + e_d$$

$$\text{Tarjonta: } Q = \beta_1 P + e_s,$$

missä Q = hyödykkeiden vaihdettu määrä, P = hinta, I = tulot, α_1 ja α_2 = kysyntäfunktion kulmakertoimet, β_1 = tarjontafunktion kulmakerroin, e_d = kysyntäfunktion virhetermi ja e_s = tarjontafunktion virhetermi. Yhtälöryhmällä on kaksi yhteistä muuttujaa: selitettävä muuttuja Q ja selittävä muuttuja P . Nämä kaksi muuttujaa ovat endogeenisia, sillä ne määräytyvät yhtälöryhmän sisällä. Kysyntäfunktiossa oleva selittävä muuttuja I sen sijaan on eksogeeninen muuttuja, sillä se tulee järjestelmän ulkopuolelta. Yhtälön oikealla puolella oleva endogeeninen muuttuja P korreloi virhetermien e_d ja e_s kanssa, minkä vuoksi pienimmän neliösumman menetelmä on harhainen ja epäkonsistentti. (Hill et al. 2012)

4.1.3 Mittausvirheet

Mallissa on mittausvirhe, jos selittävä muuttuja on mitattu virheellisesti. Vaikeasti mitattavissa oleva muuttuja on voitu korvata toisella muuttujalla, joka ei mittaa täsmällisesti haluttua asiaa. Kyseistä muuttujaa kutsutaan proxy- muuttujaksi. (Hill et al. 2012)

Mittausvirhettä voidaan mallintaa seuraavanlaisella yhtälöllä:

$$y = \beta_0 + \beta_1 x + e,$$

missä y = selitettävä muuttuja, β_0 = vakiotermi, β_1 = kulmakerroin, x = proxy-muuttuja ja e = virhetermi. Todellisen selittävän muuttujan vaikutusta selitettävään muuttujaan voidaan kuvata seuraavalla tavalla:

$$y = \beta_0 + \beta_1 x^* + v,$$

missä y = selitettävä muuttuja, β_0 = vakiotermi, β_1 = kulmakerroin, x^* = vaikeasti mitattava muuttuja ja v = virhetermi, johon sisältyy mittausvirhe. Vaikeasti mitattavan

selittävän muuttujan ja sen proxy-muuttujan välistä yhteyttä voidaan havainnollistaa seuraavasti.

$$x = x^* + u,$$

missä x = proxy-muuttuja, x^* = alkuperäinen selittävä muuttuja ja u = mittausvirhe. Tästä nähdään, että proxy-muuttuja koostuu todellisen muuttujan ja mittausvirheen summasta eikä siten mittaa täsmällisesti haluttua ominaisuutta. (Hill et al. 2012) Todellisen selittävän muuttujan vaikutus näkyy virhetermissä (Roberts & Whited, 2011). Tällöin proxy-muuttuja ja virhetermi ovat korreloituneet ja pienimmän neliösumman menetelmä ei ole konsistentti (Hill et al. 2012).

Mittausvirheen havainnollistamiseksi oletetaan, että kappaleessa 2 esitetystä esimerkissä koulutustasoa on vaikea mitata. Ongelma voidaan esittää seuraavasti:

$$\text{tulotaso} = \beta_0 + \beta_1 \text{koulutustaso} + v,$$

missä β_0 = vakiotermi, β_1 = kulmakerroin ja v = virhetermi, johon sisältyy mittausvirhe. Mallin parametrien estimoimiseksi koulutustasoa mittaavan muuttujan voisi korvata mittarilla, joka kuvaa opiskeluun käytettyjä vuosia. Tätä yhteyttä voidaan kuvata seuraavalla tavalla:

$$\text{opiskeluaika vuosina} = \text{koulutustaso} + u,$$

missä u = mittausvirhe. Yhtälöstä nähdään, että opiskeluaikaa mittaava muuttuja ei kuvaa täsmällisesti koulutustasoa, sillä siihen sisältyy mittausvirhe u . Tällöin opiskeluaikaa mittaava muuttuja on korreloitunut virhetermin kanssa ja mallissa on endogeenisuusongelma. (Hill et al. 2012) Jos malli on lineaarinen muuttujien suhteen, satunnainen mittausvirhe voidaan korjata vaihtoehtoisella instrumenttimuuttujamenetelmällä (Wang & Hsiao, 2011).

4.2 Miksi pienimmän neliösumman menetelmää ei voida käyttää?

Kun selittävä muuttuja ja virhetermi ovat korreloituneet, mallissa on endogeenisuusongelma. Tällöin pienimmän neliösumman estimointimenetelmällä ei saada konsistentteja estimaatteja. Tämä voidaan todistaa algebrallisesti seuraavalla tavalla:

- (1) Muodostetaan lineaarisen regressiomallin $y = \beta_0 + \beta_1 x + e$ ja sen systemaattisen osan $\beta_0 + \beta_1 E(x)$ erotus.

$$y - E(y) = \beta_0 + \beta_1 x + e - [\beta_0 + \beta_1 E(x)],$$

mistä saadaan:

$$y - E(y) = \beta_1 [x - E(x)] + e$$

- (2) Kerrotaan molemmat puolet tekijällä $[x - E(x)]$

$$[x - E(x)][y - E(y)] = \beta_1 [x - E(x)]^2 + [x - E(x)]e$$

- (3) Muodostetaan yhtälön kaikista tekijöistä odotusarvot:

$$E[x - E(x)][y - E(y)] = \beta_1 E[x - E(x)]^2 + E[x - E(x)]e$$

- (4) Edellinen yhtälö voidaan ilmaista myös seuraavalla tavalla:

$$\text{cov}(x, y) = \beta_1 \text{var}(x) + \text{cov}(x, e),$$

missä $\text{cov}(x, y) = E[x - E(x)][y - E(y)]$, $\text{var}(x) = E[x - E(x)]^2$ ja $\text{cov}(x, e) = E[x - E(x)]e$.

- (5) Ratkaistaan parametri β_1 :

$$\beta_1 = \frac{\text{cov}(x, y)}{\text{var}(x)} - \frac{\text{cov}(x, e)}{\text{var}(x)}$$

Pienimmän neliösumman estimaattori ilmaistaan seuraavasti:

$$b_1 = \frac{\sum(x_t - \bar{x})(y_t - \bar{y})}{\sum(x_t - \bar{x})^2} = \frac{\sum(x_t - \bar{x})(y_t - \bar{y}) / (N - 1)}{\sum(x_t - \bar{x})^2 / (N - 1)} = \frac{\hat{\text{cov}}(x, y)}{\hat{\text{var}}(x)}$$

Pienimmän neliösumman estimaattori on otoskovarianssin ja -varianssin osamäärä. Otoskoon kasvaessa otosvarianssi ja -kovarianssi konvergoituvat todelliseen va-

rianssiin ja kovarianssiin siten, että pienimmän neliösumman estimaattori b_1 konvergoituu todelliseen parametriin β_1 . Jos selittävän muuttujan ja virhetermin kovarianssi on nolla, parametri β_1 on muotoa:

$$\beta_1 = \frac{\text{cov}(x, e)}{\text{var}(x)}$$

Pienimmän neliösumman estimaatti konvergoituu parametriin β_1 :

$$b_1 = \frac{\hat{\text{cov}}(x, y)}{\hat{\text{var}}(x)} \rightarrow \frac{\text{cov}(x, y)}{\text{var}(x)} = \beta_1$$

Tällöin pienimmän neliösumman estimaattori on konsistentti. Jos selittävä muuttuja korreloi virhetermin kanssa, parametri β_1 on muotoa:

$$\beta_1 = \frac{\text{cov}(x, y)}{\text{var}(x)} - \frac{\text{cov}(x, e)}{\text{var}(x)}$$

Tällöin pienimmän neliösumman estimaattori konvergoituu parametrin β_1 ja osamäärän $\frac{\text{cov}(x, e)}{\text{var}(x)}$ summaan:

$$b_1 = \frac{\hat{\text{cov}}(x, y)}{\hat{\text{var}}(x)} \rightarrow \frac{\text{cov}(x, y)}{\text{var}(x)} = \beta_1 + \frac{\text{cov}(x, e)}{\text{var}(x)} \neq \beta_1$$

Endogeenisuusongelman tapauksessa pienimmän neliösumman estimaattoriin liittyy selittävän muuttujan ja virhetermin kovarianssin ja selittävän muuttujan varianssin osamäärän suuruinen harha ja estimaatit ovat epäkonsistentteja. Jos selittävän muuttujan ja virhetermin välinen korrelaatio on positiivista, pienimmän neliösumman menetelmä tuottaa liian suuria estimaatteja. Vastaavasti jos korrelaatio kyseisten muuttujien välillä on negatiivista, saadaan estimointimenetelmällä todellista pienemmät estimaatit. (Hill et al. 2012)

4.3 Momenttimenetelmä

Kun selittävä muuttuja on satunnainen ja korreloitunut virhetermin kanssa, pienimmän neliösumman estimaattorit ovat harhaisia ja epäkonsistentteja. Tällöin estimointiin voidaan käyttää momenttimenetelmää. Kun kaikki lineaarisen regression oletukset pitävät paikkansa, momenttimenetelmä johtaa pienimmän neliösumman es-

timaattoriin. Jos mallissa on endogeenisuusongelma, momenttimenetelmä johtaa instrumenttimuuttajamenetelmään eli kaksivaiheiseen pienimmän neliösumman menetelmään. (Hill et al. 2012)

Momenttimenetelmässä satunnaismuuttujan k:nnes momentti on k:nteen potenssiin korotettu satunnaismuuttujan odotusarvo, joka on keskiarvo äärettömistä kokeellisista tuloksista:

$$E(Y^k) = \mu_k = Y: n k: nnes populaatiomomentti,$$

missä Y = satunnaismuuttuja ja μ = keskiarvo. Vastaavasti k:nnes otosmomentti ilmaistaan seuraavasti:

$$\widehat{E(Y^k)} = \hat{\mu}_k = Y: n k: nnes otosmomentti = \sum_{t=1}^T y_t^k / N$$

Momenttimenetelmässä m kappaletta populaatiomomentteja asetetaan yhtä suuriksi m kappaleen otosmomentin kanssa ja tällöin voidaan estimoida m kappaletta tuntemattomia parametreja. (Hill et al. 2012)

Satunnaismuuttujan Y odotusarvo $E(Y)$ ja varianssi $\text{var}(Y)$ ilmaistaan seuraavasti:

$$E(Y) = \mu$$

$$\text{var}(Y) = \sigma^2 = E(Y - \mu)^2 = E(Y^2) - \mu^2$$

Jotta voidaan estimoida populaation keskiarvo, asetetaan ensimmäinen populaatiomomentti ja ensimmäinen otosmomentti yhtä suuriksi:

$$\mu_1 = \sum y_i / N = \bar{y},$$

missä μ_1 = ensimmäinen populaatiomomentti, $\sum y_i / N$ = ensimmäinen otosmomentti ja \bar{y} = otoskeskiarvo. (Hill et al. 2012)

Populaation varianssin estimoimiseksi korvataan toinen populaatiomomentti sen otosarvolla ja ensimmäinen momentti otoskeskiarvolla:

$$\text{var}(Y) = \tilde{\sigma}^2 = \hat{\mu}_2 - \hat{\mu}^2 = \frac{\sum y_i^2}{N} - \bar{y}^2 = \frac{\sum y_i^2 - N\bar{y}^2}{N} = \frac{\sum (y_i - \bar{y})^2}{N},$$

missä $\hat{\mu}_2 = \frac{\sum y_i^2}{N}$ = toinen otosmomentti, $\hat{\mu}^2 =$ neliöön korotettu ensimmäinen otosmomentti, \bar{y} = otoskeskiarvo. Populaation keskiarvon estimaatiksi saatiin otoskeskiarvo \bar{y} ja varianssin estimaatiksi otosvariassi $\frac{\sum (y_i - \bar{y})^2}{N}$. Momenttimenetelmällä estimoidut estimaattorit ovat konsistentteja ja konvergoituvat todellisiin parametriarvoihin suurissa otoksissa. (Hill et al. 2012)

Lineaarisen regression oletuksista voidaan muotoilla momenttiehdot:

$$(1) E(e) = 0$$

$$(2) E(xe) = 0$$

Kun kaksi populaatiomomenttia korvataan kahdella otosmomentilla, saadaan kaksi yhtälöä, joissa on kaksi tuntematonta parametria:

$$\frac{1}{N} \sum (y_i - b_0 - b_1 x_i) = 0$$

$$\frac{1}{N} \sum x_i (y_i - b_0 - b_1 x_i) = 0$$

Kun nämä yhtälöt ratkaistaan, saadaan pienimmän neliösumman estimaattorit b_0 ja b_1 , jotka on esitetty kappaleessa 3. Jos mallissa on endogeenisuusongelma, momenttimenetelmä johtaa instrumenttimuuttujaeestimaattoreihin. (Hill et al. 2012)

4.3.1 Instrumenttimuuttujamenetelmä eli kaksivaiheinen pienimmän neliösumman menetelmä

Toinen momenttiehto $E(xe)=0$ esittää, että selittävä muuttuja ei saa korreloida virhetermin kanssa. Tällöin pienimmän neliösumman estimaattori ei ole konsistentti. Kuitenkin jos on olemassa jokin muuttuja z , jolla on tietyt ominaisuudet, voidaan momenttimenetelmän avulla saada konsistentteja estimaatteja. Muuttujaa z kutsutaan instrumenttimuuttujaksi, sillä se toimii eräänlaisena instrumenttina mallissa. Instrumenttimuuttujalla ei saa olla suoraa vaikutusta selitettävään muuttujaan eikä se saa korreloida virhetermin kanssa eli sen tulee olla eksogeeninen. Lisäksi kyseisen muut-

tujan on korreloitava riittävän vahvasti endogeenisen muuttujan kanssa. Instrumenttimuuttujan avulla voidaan muodostaa kolmas momenttiehto:

$$(3) E(ze) = 0,$$

mistä ilmenee, että instrumenttimuuttuja ja virhetermi eivät saa korreloida keskenään. Ensimmäisen ja kolmannen momenttiehdon avulla saadaan estimaatit parametreille β_0 ja β_1 . Korvaamalla kaksi momenttipopulaatiota kahdella otosmomentilla, saadaan kaksi yhtälöä:

$$\frac{1}{T} \sum (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0$$

$$\frac{1}{T} \sum z_i (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0$$

Näiden yhtälöiden ratkaisu johtaa momenttimenetelmän estimaattoreihin, joita kutsutaan instrumenttimuuttujaestimaattoreiksi:

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

$$\hat{\beta}_1 = \frac{\sum (z_i - \bar{z})(y_i - \bar{y})}{\sum (z_i - \bar{z})(x_i - \bar{x})}$$

Estimoitu vakiotermin $\hat{\beta}_0$ saadaan pienimmän neliösumman menetelmän tavoin selitettävän muuttujan otoskeskiarvoa selittävästä yhtälöstä. Kulmakertoimen $\hat{\beta}_1$ instrumenttimuuttujaestimaattori eroaa pienimmän neliösumman vastaavasta estimaattorista siten, että yhtälön osoittaja esittää selitettävän muuttujan ja instrumenttimuuttujan hajonnan yhteisvaihtelua ja nimittäjä selitettävän muuttujan ja instrumenttimuuttujan yhteisvaihtelua. Parametrit $\hat{\beta}_0$ ja $\hat{\beta}_1$ ovat konsistentteja, jos instrumenttimuuttujalle asetetut ehdot täyttyvät. (Hill et al. 2012)

Usean muuttujan regressiomallissa instrumenttimuuttujaestimointi toteutetaan kahdessa vaiheessa:

- (1) Oletetaan, että yhtälössä $y = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + e$, muuttuja x_k on endogeeninen ja muuttujat (x_1, \dots, x_{k-1}) ovat eksogeenisiä. Endogeenisen muuttujan instrumenttimuuttujaa kuvataan symbolilla z_1 . Ensimmäisen vaiheen yhtälössä

selitetään endogeenista muuttujaa sen instrumenttimuuttujalla ja kaikilla eksogeenisillä muuttujilla:

$$x_k = \gamma_0 + \gamma_1 x_1 + \dots + \gamma_{k-1} x_{k-1} + \theta_1 z_1 + v,$$

missä x_k = endogeeninen muuttuja, γ_0 = vakiotermin estimaatti, (x_1, \dots, x_{k-1}) = selittävät eksogeeniset muuttujat, $(\gamma_1, \dots, \gamma_{k-1})$ = selittävien eksogeenisten muuttujien parametriesiimaatit, z_1 = instrumenttimuuttuja, θ_1 = instrumenttimuuttujan kerroin ja v = ensimmäisen vaiheen virhetermi. Yhtälö muodostetaan jokaiselle endogeeniselle muuttujalle erikseen. Koska kaikki yllä olevan yhtälön oikean puolen selittävästä muuttujista on eksogeenisia, yhtälö voidaan estimoida pienimmän neliösumman menetelmällä. Tällöin saadaan endogeenisen selittävän muuttujan estimoitu yhtälö:

$$\hat{x}_k = \hat{\gamma}_0 + \hat{\gamma}_1 x_1 + \dots + \hat{\gamma}_{k-1} x_{k-1} + \hat{\theta}_1 z_1$$

(2) Toisessa vaiheessa korvataan alkuperäisen yhtälön endogeeninen muuttujaa sen estimaatilla:

$$y = \beta_0 + \beta_1 \hat{x}_k + e^*,$$

missä y = selitettävä muuttuja, β_0 = vakiotermi, β_1 = kulmakerroin, \hat{x}_k = estimoitu selittävä muuttuja ja e^* = toisen vaiheen virhetermi. Yhtälöstä saadaan estimaattorit $\hat{\beta}_0$ ja $\hat{\beta}_1$, joita kutsutaan instrumenttimuuttujaestimaattoreiksi eli kaksivaiheisen pienimmän neliösumman estimaattoreiksi. (Hill et al. 2012)

4.3.2 Instrumenttimuuttujan pätevyden testaaminen

Jos instrumenttimuuttuja korreloi virhetermin kanssa, kolmas momenttiehto $E(ze) = 0$ ei täyty. Tällöin instrumenttimuuttuja ei ole pätevä eikä instrumenttimuuttujaestimaattorit ole konsistentteja. Instrumenttimuuttujien pätevyyttä voidaan arvioida, jos instrumenttimuuttujia on vähintään yhtä monta kuin mahdollisia endogeenisia muuttujia.

Usean selittävän muuttujan regressiomallissa instrumenttimuuttujan pätevyyttä voidaan testata seuraavalla tavalla:

(1) Muodostetaan yhtälö siten, että selittävinä muuttujina ovat instrumenttimuuttujat ja eksogeeniset muuttujat:

$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + \theta_1 z_1 + \dots + \theta_L z_L + v,$$

missä y = selitettävä muuttuja, β_0 = vakiotermi, (x_1, \dots, x_k) = eksogeeniset selittävät muuttujat, $(\beta_1, \dots, \beta_k)$ = eksogeenisten muuttujien parametrit, (z_1, \dots, z_k) = instrumenttimuuttujat, $(\theta_1, \dots, \theta_L)$ = instrumenttimuuttujien parametrit ja v = virhetermi. (Roberts & Whited, 2011)

(2) Yhtälöstä saadaan residuaalit $\hat{e} = y - \hat{\beta}_0 - \hat{\beta}_1 x_1, \dots, \hat{\beta}_k x_k$.

(3) Muodostetaan jokaiselle instrumenttimuuttujalle yhtälö, siten että selittävänä muuttujana on edellisessä vaiheessa saatu residuaali.

(4) Arvioidaan instrumenttimuuttujan pätevyyttä otoskoolla kerrotun selitysasteen avulla. (Hill et al. 2012)

Empiirisessä tutkimuksessa instrumenttimuuttujamenetelmässä käytettyjen instrumenttimuuttujien ja endogeenisten muuttujien välinen korrelaatio on usein heikkoa. (Staiger & Stock, 1997). Tällöin saaduilla estimaateilla on suuret keskivirheet. Instrumenttimuuttujien käytössä voidaan havaita kaksi perusongelmaa. Ensiksi, jos korrelaatio instrumenttimuuttujan ja endogeenisen selittävän muuttujan välillä on heikkoa, instrumenttimuuttujaestimaattorit voivat olla epäkonsistentteja. Toiseksi äärellisissä otoksissa instrumenttimuuttujaestimaattorit ovat pienimmän neliösumman estimaattoreiden tapaan harhaisia, kun instrumenttien ja endogeenisten muuttujien välinen selityskerroin lähenee nollaa. (Bound, Jaeger & Baker, 1995)

4.4 Endogeenisuuden testaaminen

Jos selittävän muuttujan ja virhetermin välillä on korrelaatiota, ei pienimmän neliösumman menetelmää voida käyttää. Pienimmän neliösumman menetelmälle vaihtoehtoinen estimointimenetelmä on instrumenttimuuttujamenetelmä. Jotta tiedetään, mitä estimointimenetelmää voidaan käyttää, tulee selittävän muuttujan ja virhetermin välinen korrelaatio selvittää. Tämä voidaan toteuttaa Hausman- testin avulla. Testin nollahypoteesina, että selittävä muuttuja ja virhetermi eivät korreloi keskenään. Hausman-testi vertaa pienimmän neliösumman menetelmän ja instrumenttimuuttujamenetelmän toimivuutta.

Hausman- testi toteutetaan seuraavien vaiheiden mukaisesti:

- (1) Ensimmäisessä vaiheessa selitetään endogeenista muuttujaa kaikilla mallin eksogeenisilla muuttujilla ja instrumenttimuuttujilla:

$$x = \gamma_0 + \theta_1 z_1 + \dots + \theta_L z_L + v,$$

missä x = endogeeninen selittävä muuttuja, γ_0 = vakiotermi, (z_1, \dots, z_L) = instrumenttimuuttujat, $(\theta_1, \dots, \theta_L)$ = instrumenttimuuttujien kertoimet ja v = virhetermi. Yhtälö muodostetaan jokaiselle endogeeniselle muuttujalle erikseen. Ensimmäisestä vaiheesta saadaan residuaalit:

$$\hat{v} = x - \hat{\gamma}_1 - \hat{\theta}_1 z_1 - \dots - \hat{\theta}_L z_L$$

- (2) Ensimmäisen vaiheen residuaalit lisätään alkuperäiseen yhtälöön:

$$y = \beta_0 + \beta_1 x + k\hat{v} + e,$$

missä y = selitettävä muuttuja, β_0 = vakiotermi, β_1 = kulmakerroin, x = selittävä muuttuja, \hat{v} = ensimmäisen vaiheen residuaali, k = edellisen kerroin ja e = virhetermi. Kyseinen yhtälö estimoidaan pienimmän neliösumman menetelmällä ja Hausman-testin nollahypoteesin paikkansapitävyys ratkaistaan t-testin avulla.

Jos kertoimen k estimaatti ei t-testin mukaan poikkea tilastollisesti merkittävästi nollasta, Hausman-testin nollahypoteesi jää voimaan. Tällöin selittävän muuttujan ja virhetermin välillä ei ole korrelaatiota. Hausman-testin jäädessä voimaan sekä pienimmän neliösumman estimaattori että instrumenttimuuttujaestimaattori ovat konsistentteja. Näiden kahden estimaattorin ero konvergoituu nollaan suurissa otoksissa. Tällöin tulee käyttää tehokkaampaa pienimmän neliösumman estimaattoria.

Vastaavasti, jos kertoimen k estimaatti poikkeaa tilastollisesti merkitsevästi nollasta, Hausman-testin nollahypoteesi hylätään ja selittävän muuttujan ja virhetermin välillä voidaan olettaa olevan korrelaatiota. Tämä johtaa siihen, että pienimmän neliösumman menetelmä ei ole konsistentti ja on käytettävä instrumenttimuuttujaestimointimenetelmää. (Hill et al. 2012)

5 Esimerkki endogeenisuusongelmasta: tutkimus siirtomaiden instituutioiden vaikutuksesta maiden taloudelliseen suorituskykyyn

5.1 Tutkimuksen taustaa

Acemoglu, Johnson ja Robinson (2000) ovat tutkineet siirtomaihin perustettujen instituutioiden vaikutusta maiden taloudelliseen suorituskykyyn. Eurooppalaisten kolonisaatiostrategioiden välillä oli suuria eroja, ja siten eri maihin syntyi erilaisia instituutioita. Kolonialismin aikana syntyneet siirtomaiden olosuhteet ja instituutiot ovat säilyneet maiden itsenäisyyden jälkeenkin. Paremmat instituutiot mahdollistavat investoimaan enemmän fyysiseen ja henkiseen pääomaan, mikä johtaa parempaan taloudelliseen suorituskykyyn. (Acemoglu, Johnson & Robinson, 2000)

Ääritapauksena eurooppalaiset loivat eristettyjä valtioita, joiden instituutiot eivät toimineet yksityisen omistuksen puolesta eivätkä valvoneet julkisen vallan pakkolunastusta. Eristyneiden valtioiden päätarkoitus oli siirtää siirtomaiden resursseja eurooppalaisille kolonialisteille. Toisena ääritapauksena luotiin ”uuseurooppalais-” valtioita, joissa uudisasukkaat kopioivat eurooppalaisia instituutioita korostaen yksityisomistusta ja julkisen vallan valvontaa. (Acemoglu et al. 2000)

Kolonisaatiostrategiaan vaikutti keskeisesti kysymys siitä, pystyivätkö eurooppalaiset asettumaan siirtomaahan. Eurooppalaiset eivät voineet asettua maihin, joissa kuolleisuus oli suurta, jolloin todennäköisyys eristettyjen valtioiden luontiin oli suuri. (Acemoglu et al. 2000)

5.2 Tutkimuksen hypoteesit

Tutkimuksen hypoteesit voidaan esittää seuraavien yhtälöiden avulla:

1. Siirtomaiden nykyiset instituutiot vaikuttavat maiden taloudelliseen suorituskykyyn:

$$\log y_i = \mu + \alpha R_i + X_i' \gamma + e_i,$$

missä $\log y_i$ = asukasta kohti laskettu tulotaso, μ = vakiotermi, R_i = nykyiset instituutiot, X_i' = vektori, joka kuvaa muita selittäviä muuttujia, γ = edellisen kerroin ja e_i = virhetermi.

2. Kolonialismin aikana syntyneet instituutiot ja nykyiset instituutiot korreloivat keskenään:

$$R_i = \lambda_R + \beta_R C_i + X_i' \gamma_R + v_{Ri},$$

missä R_i = nykyiset instituutiot, λ_R = vakiotermi, β_R = kulmakerroin, C_i = kolonialismin aikana syntyneet instituutiot, X_i' = vektori, joka kuvaa muita selittäviä muuttujia, γ_R = edellisen kerroin ja v_{Ri} = virhetermi.

3. Kolonisaatiostrategia riippui siitä, pystyivätkö eurooppalaiset asettumaan siirtomaahan:

$$C_i = \lambda_C + \beta_C S_i + X_i' \gamma_C + v_{Ci},$$

missä C_i = kolonialismin aikana syntyneet instituutiot, λ_C = vakiotermi, β_C = kulmakerroin, S_i = kolonialistien asettuminen siirtomaahan, X_i' = vektori, joka kuvaa muita selittäviä muuttujia, γ_C = edellisen kerroin ja v_{Ci} = virhetermi.

4. Eurooppalaisten kolonialistien kohtaamat siirtomaiden kuolleisuusluvut vaikuttivat kyseiseen maahan asettumiseen:

$$S_i = \lambda_S + \beta_S \log M_i + X_i' \gamma_S + v_{Si},$$

missä S_i = kolonialistien asettuminen siirtomaahan, λ_S = vakiotermi, β_S = kulmakerroin, $\log M_i$ = kolonialistien kohtaamat kuolleisuusluvut, X_i' = vektori, joka kuvaa muita selittäviä muuttujia, γ_S = edellisen kerroin ja v_{Si} = virhetermi.

5.3 Tutkimuksen aineisto ja muuttajat

Tutkimuksen otos koostuu 71 maasta, jotka ovat olleet eurooppalaisten maiden siirtomaita 1900-luvulla. Kyseiset maat ja niiden eri muuttajien arvot on esitetty liitteen 1 taulukossa.

Taloudellisen suorituskyvyn mittarina on käytetty dollarimääräistä asukasta kohdin laskettua bruttokansantuotetta vuonna 1995. Kyseinen mittari on peräisin World DataBank – tietokannasta. Useiden muiden taloudellisten muuttajien tavoin, bruttokansantuotteen jakauma on vino. Jotta jakauma olisi lähempänä normaalijakaumaa, muuttujasta on tehty logaritmuunnos. (Hill et al. 2012)

Nykyisiä instituutioita on kuvattu indeksillä, joka mittaa julkisen vallan pakkolunastuksen riskiltä suojautumista asteikolla 0-10. Maat, joissa kyseiseltä riskiltä suojautuminen oli heikointa, saavat arvon nolla ja vastaavasti maat, joissa riskiltä suojautuminen oli vahvinta, saavat arvon 10. Odotuksena on, että kolonialistien luomat eristetyt valtiot saivat arvon nolla ja ”uuseurooppalais”-valtiot arvon 10. (Acemoglu et al., 2000)

Kolonialismin aikana syntyneitä instituutioita on mitattu indeksillä, joka kuvaa toimeenpanovallan rajoittamista 1900-luvulla. Muuttuja saa arvoja väliltä 1-7. Mikäli maa ei ollut itsenäinen kyseisenä ajanjaksona, se saa arvon yksi.

Eurooppalaisten kolonialistien asettumista siirtomaahan on kuvattu mittarilla, joka kertoo eurooppalaista alkuperää olevien henkilöiden prosenttiosuuden väestöstä 1900-luvulla.

Kuolleisuuslukuja mittaava muuttuja kuvaa estimoituja kuolleisuuslukuja antaen maalle arvon väliltä 1,7 – 6,2. Muuttujasta tehdään logaritmuunnos, jotta Afrikan maiden korkeilla kuolleisuusluvuilla ei ole liian suurta vaikutusta tulokseen (Acemoglu et al. 2000).

Taloudellista suorituskykyä mittaavaa muuttujaa lukuun ottamatta aineistot ovat peräisin MIT Economics – sivustolta.

5.4 Tutkimuksessa ilmenevät ongelmat

5.4.1 Mallin lineaarisuus

Instituutioita mittaavan indeksin ja asukasta kohdin lasketun bruttokansantuotteen välistä lineaarisuutta voidaan tutkia muodostamalla instituutioita mittaavan indeksin jakaumasta dummy-muuttujat D_1 , D_2 , D_3 ja D_4 :

D_1 = indeksin arvot pienempiä tai yhtä suuria kuin 5,125

D_2 = indeksin arvot välillä (5,126; 6,75)

D_3 = indeksin arvot välillä (6,76; 8,375)

D_4 = indeksin arvot suurempia tai yhtä suuria kuin 8,376

(Acemoglu et al. 2000)

Kun haluttu ominaisuus on läsnä, dummy-muuttuja saa arvon yksi ja vastaavasti kun ominaisuus ei ole läsnä, se saa arvon nolla. (Hill et al. 2012)

Muuttujien lineaarisuuden selvittämiseksi muodostetaan regressiomalli siten, että selitettävä muuttuja on taloudellista suorituskyykyä mittaava muuttuja ja selittävinä muuttujina ovat dummy-muuttujat:

$$\log y_i = \beta_0 + \beta_1 R_i + \gamma_1 (R_i * D_1) + \gamma_2 (R_i * D_2) + \gamma_3 (R_i * D_3) + \gamma_4 (R_i * D_4) + e,$$

missä $\log y_i$ = selitettävä muuttuja, β_0 = vakiotermi, R_i = nykyiset instituutiot, β_1 = nykyisten instituutioiden kulmakerroin, D_1 = ensimmäisen kvartaalin dummy-muuttuja, D_2 = toisen kvartaalin dummy-muuttuja, D_3 = kolmannen kvartaalin dummy-muuttuja, D_4 = neljännen kvartaalin dummy-muuttuja, $(\gamma_1, \dots, \gamma_4)$ = kulmakertoimet ja e = virhetermi. Muuttuja $(R_i * D_k)$ on interaktiomuuttuja, joka kuvaa instituutioiden ja eri kvartaalien yhteisvaikutusta. (Hill et al. 2012) Vertailuryhmänä on ensimmäiseen kvartaaliin kuuluvat maat eli maat, joilla on suurin julkisen vallan pakkolunastuksen riski. Hypoteesina on, että indeksin arvon kasvaessa, vaikutus taloudelliseen suorituskyykyyn on positiivisesti suurempi. (Acemoglu et al. 2000)

5.4.2 Puuttuvat selittävät muuttujat

Tutkimus tulee ongelmalliseksi, jos mallista puuttuu oleellisia selittäviä muuttujia, jotka ovat korreloituneet instituutioita mittaavan muuttujan kanssa. Instituutioita mittaavan muuttujan ja potentiaalisten mukaan otettavien muuttujien välistä korrelaatiota voidaan tutkia seuraavan yhtälön avulla:

$$y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_3x_3 + \beta_4x_4 + \beta_5x_5 + \beta_6x_6 + v$$

missä y = selitettävä muuttuja, β_0 = vakiotermi, x_1 = nykyiset instituutiot, x_2 = ilmasto, x_3 = maantiede, x_4 = uskonto, x_5 = luonnonvarat, x_6 = maaperän laatu, $(\beta_1, \dots, \beta_6)$ = kulmakertoimet ja v = yhdistetty virhetermi. (Acemoglu et al. 2000) Jos yhdistetty virhetermi korreloi jonkun havaitsemattoman selittävän muuttujan kanssa, instituutioita mittaava muuttuja on endogeeninen (Roberts & Whited, 2011).

5.4.3 Samanaikaisuusharha

Acemoglu et al. (2000) esittävät, että koska instituutioiden mittaus perustuu nykyisiin instituutioihin, on todennäköistä että rikkailla talouksilla on varaa parempiin instituutioihin. Tällöin kausaalisuuden suunta on epäselvä: ei voida täsmällisesti sanoa vaikuttavatko instituutiot taloudelliseen suorituskyykyyn vai taloudellinen suorituskyyky instituutioihin. Tässä tutkielmassa ei keskitytä puuttuvien selittävien muuttujien aiheuttamaan endogeenisuusongelmaan eikä samanaikaisuusharhaan, sillä Acemoglu et al. (2000) toteavat että suorittamassaan tutkimuksessa keskeisin endogeenisuutta aiheuttava ongelma on mittausvirhe.

5.4.4 Mittausvirhe

Instituutiot on mitattu jälkikäteen ja sen vuoksi mittaukseen voi liittyä harha siitä, että rikkaissa talouksissa oletetaan olevan parempia instituutioita. Instituutioiden mittarina on käytetty indeksiä, joka mittaa julkisen vallan pakkolunastuksen riskiltä suojautumista asteikolla 0-10. (Acemoglu et al., 2000) Kyseinen indeksi ei kuvaa nykyisiä ins-

tituutioita täsmällisesti eli se on instituutioiden proxy-muuttuja. Tällöin instituutioita mittaava muuttuja on korreloitunut virhetermin kanssa. (Hill et al. 2000)

5.5 Endogeenisuuden korjaus

Jos mallissa on endogeenisuusongelma, se voidaan korjata instrumenttimuuttujan avulla. Acemoglu et al. (2000) esittävät, että eurooppalaisten kolonialistien kohtaamisen kuolleisuuslukujen muutokset aiheuttavat mahdollisesti instituutioissa eksogeenistä vaihtelua. Tämän perusteella kuolleisuusluvuista muodostetaan instrumenttimuuttuja instituutioille. Kuolleisuuslukujen ja instituutioiden välinen yhteys esitetään seuraavasti:

$$R_i = \zeta + \beta \log M_i + X_i' \delta + v_i$$

missä R_i = instituutioita mittaava indeksi, ζ = vakiotermin, β = kulmakerroin, $\log M_i$ = kuolleisuuslukujen logaritmi, X_i' = vektori, joka kuvaa muita selittäviä muuttujia, δ = edellisen kerroin ja v_i = virhetermi. Tämä instrumenttistrategia toimii vain, jos instrumenttimuuttuja $\log M_i$ täyttää kappaleessa 4.3.1. määritetyt ehdot. Kun kuolleisuusluvuista otetaan logaritmi, kuolleisuuslukujen ja instituutioita mittaavan muuttujan välillä on suhteellisen lineaarinen yhteys. (Acemoglu et al. 2000)

Estimoidaan alkuperäinen yhtälö ja yllä mainittu yhtälö instrumenttimuuttujamenetelmällä kahdessa vaiheessa:

- (1) Ensimmäisessä vaiheessa selitetään endogeenista muuttujaa R_i sen instrumenttimuuttujalla $\log M_i$ ja eksogeenisilla muuttujilla:

$$R_i = \zeta + \beta \log M_i + X_i' \delta$$

Yhtälöstä saadaan endogeenisen muuttujan estimaatti \hat{R}_i .

- (2) Sijoitetaan endogeenisen muuttujan estimaatti alkuperäiseen yhtälöön.

$$\log y_i = \mu + \alpha \hat{R}_i + X_i' \gamma + e$$

Toisesta vaiheesta saadaan parametrien μ ja α instrumenttimuuttujaestimaattorit:

$$\hat{\mu} = \log \bar{y} - \hat{\alpha} \bar{R}$$

$$\hat{\alpha}_1 = \frac{\sum(\log M_i - \log \bar{M})(\log y_i - \log \bar{y})}{\sum(\log M_i - \log \bar{M})(R_i - \bar{R})}$$

5.6 Endogeenisuuden testaaminen

Mallin endogeenisuutta voidaan testata Hausman-testin avulla. Tällöin nollahypoteesina on, että instituutioita mittaava selittävä muuttuja ei korreloi virhetermin kanssa.

Hausman-testi toteutetaan kahdessa vaiheessa:

(1) Ensimmäisen vaiheen yhtälö on muotoa:

$$R_i = \zeta + \beta \log M_i + X_i' \gamma + v,$$

missä R_i = nykyiset instituutiot, ζ = vakiotermi, $\log M_i$ = kolonialistien kohtaamat kuolleisuusluvut, β = edellisen kerroin, X_i' = vektori, joka kuvaa muita selittäviä muuttujia, γ = edellisen kerroin ja v = virhetermi. Ensimmäisen vaiheesta saadaan residuaaliksi:

$$\hat{v} = R_i - \hat{\zeta} - \hat{\theta}_1 \log M_i - X_i' \gamma$$

(2) Toiseen vaiheen yhtälö muodostetaan seuraavasti:

$$\log y_i = \mu + \beta R_i + k \hat{v} + e$$

Hausman-testin nollahypoteesin paikkansapitävyys testataan t-testillä. Jos kerroin k on tilastollisesti merkitsevä, voidaan olettaa, että instituutioita mittaava muuttuja on endogeeninen. Tällöin estimointi suoritetaan instrumenttimuuttujamenetelmän avulla.

6 Tutkimustulokset

Tämän kappaleen taulukoihin on koottu kappaleessa 5.2. esitettyjen hypoteesien mukaisten testien tulokset. Tulokseni poikkeavat osittain alkuperäisen tutkimuksen tuloksista, sillä käytössäni oleva aineisto on hieman erilainen. Taulukossa esitetään muuttujien osalta kaksi lukua: ensimmäinen kuvaa parametriestimaatin arvoa ja su-

luissa oleva luku on parametrin keskivirhe. Selitysaste kuvaa sitä, kuinka paljon selittävä muuttuja kertoo selitettävän muuttujan vaihtelusta. (Hill et al. 2012)

TAULUKKO 1. Pienimmän neliösumman estimaatit, Hausman-testi ja instrumenttimuuttujaestimaatit hypoteesille 1

	1	2	3	4	5
Vakiotermi	4,76 (0,41)	7,35 (0,19)	8,04 (0,12)	9,14 (0,61)	1,78 (1,09)
Julkisen vallan pakkolunastuksen riskiltä suojautuminen	0,51 (0,06)				0,97 (0,17)
Julkisen vallan pakkolunastuksen riskiltä suojautumisen dummy toisessa kvartaalissa		0,30 (0,24)			
Julkisen vallan pakkolunastuksen riskiltä suojautumisen dummy kolmannessa kvartaalissa		1,19 (0,25)			
Julkisen vallan pakkolunastuksen riskiltä suojautumisen dummy neljännessä kvartaalissa		2,65 (0,39)			
Kolonialistien kohtaamat kuolleisuusluvut (log)				-0,58 (0,13)	
Residuaali_Julkisen vallan pakkolunastuksen riskiltä suojautuminen			0,36 (0,09)		
Korjattu selitysaste (Adjusted R²)	0,52	0,47	0,18	0,24	0,34

Taulukon 1 ensimmäisessä sarakkeessa esitetään taloudellisen suorituskyvyn ja nykyisten instituutioiden välistä yhteyttä kuvaavan hypoteesin tulokset. Estimointiin on käytetty pienimmän neliösumman menetelmää. Sekä vakiotermi että kulmakerroin

ovat tilastollisesti merkitseviä. Selitysaste kertoo, että julkisen vallan pakkolunastuksen riskiltä suojautumista kuvaava muuttuja selittää noin puolet taloudellista suorituskykyä mittaavasta muuttujasta.

Toisessa sarakkeessa on kuvattu edellä mainittujen muuttujien lineaarista yhteyttä tarkastelevan hypoteesin tulokset. Taulukosta nähdään toiseen, kolmanteen ja neljänteen kvartaaliin kuuluvien maiden ero ensimmäiseen kvartaaliin kuuluviin maihin nähden. Taloudellisen suorituskyvyn ja instituutioiden välinen yhteys voidaan olettaa lineaariseksi, sillä indikaattorimuuttujat saavat kertoimet odotetussa järjestyksessä: mitä voimakkaampaa suojautuminen julkisen vallan pakkolunastuksen riskiä vastaan on, sitä suurempi positiivinen vaikutus sillä on taloudelliseen suorituskykyyn.

Koska instituutioiden mittarina on käytetty proxy-muuttujaa, on todennäköistä että mallissa on endogeenisuusongelma. Tämän selvittämiseksi olen suorittanut Hausman-testin, vaikka kyseistä testiä ei alkuperäisessä tutkimuksessa ole suoritettu. Hausman-testin tulokset olen koonnut kolmanteen sarakkeeseen. Hausman-testin nollihypoteesi hylätään, sillä residuaalin kerroin poikkeaa tilastollisesti merkittävästi nollostä. Tämä tarkoittaa, että instituutioita mittaava muuttuja on endogeeninen ja korreloitunut mallin virhetermin kanssa. Tällöin ensimmäisessä sarakkeessa kuvatut pienimmän neliösumman estimaatit eivät ole luotettavia.

Koska instituutioita mittaavan muuttujan ja taloudellisen suorituskyvyn välinen yhteys on lineaarinen, voidaan endogeenisuus korjata instrumenttimuuttujan avulla (Wang & Hsiao, 2011). Tutkimuksessa on käytetty instrumenttimuuttujana kolonialistien kohtaamia kuolleisuuslukuja. Kaksivaiheisen pienimmän neliösumman menetelmän ensimmäisen vaiheen tulokset ovat sarakkeessa 4 ja toisen vaiheen tulokset sarakkeessa 5. Tulokset esittävät, että instituutioita mittaavan muuttujan vaikutus taloudelliseen suorituskykyyn on hieman suurempi kuin pienimmän neliösumman menetelmällä estimoitaessa. Tuloksista havaitaan myös, että selitysaste on pienempi kuin sarakkeessa 1 esitetty selitysaste. Tämä tarkoittaa, että instituutioiden vaihtelun vaikutus taloudelliseen suorituskykyyn oli arvioitu liian suureksi pienimmän neliösumman menetelmällä estimoitaessa.

TAULUKKO 2. Pienimmän neliösumman estimaatit hypoteeseille 2, 3 ja 4.

Selitettävä muuttuja: julkisen vallan pakkolunastuksen riskiltä suojautuminen	
Vakiotermi	5,66 (0,25)
Kolonialismin aikana syntyneet instituutiot	0,33 (0,08)
Korjattu selitysaste (Adjusted R²)	0,21
Selitettävä muuttuja: kolonialismin aikana syntyneet instituutiot	
Vakiotermi	1,27 (0,22)
Kolonialistien asettuminen siirtomaahan	0,05 (0,01)
Korjattu selitysaste (Adjusted R²)	0,45
Selitettävä muuttuja: Kolonialistien asettuminen siirtomaahan	
Vakiotermi	63,62 (10,06)
Kolonialistien kohtaamat kuolleisuusluvut	-10,42 (2,08)
Korjattu selitysaste (Adjusted R²)	0,26

Taulukossa 2 on esitetty muiden kuin taloudellisen suorituskyvyn ja nykyisten instituutioiden välistä yhteyttä kuvaavien hypoteesien tulokset. Taulukon ensimmäisessä osiossa on kuvattu kolonialismin aikaisten ja nykyisten instituutioiden välistä korrelaatiota. Keskimmäisessä osassa on tulokset hypoteesille, joka esittää kolonialismin ai-

kaisten siirtomaiden riippuvuutta siitä, pystyttiinkö kyseiseen maahan asettumaan. Viimeinen osio esittää kolonialistien kohtaamien kuolleisuuslukujen vaikutuksen siirtomaahan asettumiseen. Tuloksien perusteella voidaan päätellä, että minkään hypoteesin selittävä muuttuja ei selitä kovinkaan paljoa selitettävän muuttujan vaihtelusta, mutta muuttujat korreloivat keskenään. Tämän tutkielman pääpaino on taloudellisen suorituskyvyn ja siirtomaiden nykyisten instituutioiden välisessä suhteessa, mutta sitä voisi laajentaa tarkastelemaan myös muiden muuttujien välisiä yhteyksiä.

7 Johtopäätökset

Endogeenisuusongelma syntyy, kun mallin selittävä muuttuja on korreloitunut virhetermin kanssa. Tällöin pienimmän neliösumman estimointimenetelmän avulla ei saada luotettavia tuloksia. Tämän vuoksi endogeenisuusongelmaa aiheuttavat tilanteet on tärkeä tuntea. Mittausvirhe kuvaa virheellisesti mitatun selittävän muuttujan ja virhetermin välistä korrelaatiota. Jos mallista puuttuu oleellisia selittäviä muuttujia, malliin mukaan otettu selittävä muuttuja on endogeeninen jos se korreloi kyseisten puuttuvien selittävien muuttujien kanssa. Samanaikaisuusharhaa esiintyy moniyhtälömallissa, kun kausaalisuuden suunta ei ole selvä. Tällöin mallin sisällä määräytyvät endogeeniset muuttujat ovat korreloituneet yhtälöiden virhetermien kanssa. Kaikki nämä ongelmat voidaan ratkaista instrumenttimuuttujan avulla.

Endogeenisuusongelmaa esiintyy hyvin arkipäiväisissä tilanteissa. Kappaleessa 4.1.3. käsiteltiin esimerkkiä, jossa henkilön koulutustason ja tulotason välillä havaittiin endogeenisuusongelma. Koulutustasoa mittaava muuttuja oli korvattu muuttujalla, joka mittaa koulutukseen käytettyjä vuosia, jolloin mallissa oli mittausvirhe.

Endogeenisuusongelmaa käsiteltiin myös tarkastelemalla siirtomaiden instituutioiden ja taloudellisen suorituskyvyn välistä suhdetta, jota Acemoglu et al. (2000) ovat tutkineet. Kuitenkin suuri osa analysoimani tutkimuksen sisällöstä jäi käsittelemättä. Tutkielmani pääpaino oli mittausvirheen aiheuttamassa endogeenisuusongelmassa. Sitä voisi laajentaa tarkastelemalla myös puuttuvien selittävien muuttujien aiheuttamaa endogeenisuusongelmaa sekä samanaikaisuusharhaa. Mallin identifioitavuutta voi-

taisiin tarkastella vaihtoehtoisten aineistojen avulla, kuten Acemoglu et al. (2000) ovat tutkimuksessaan tehneet. Lisäksi tutkimuksen analysointia voisi kehittää edelleen siten, että myös muiden kuin taloudellisen suorituskyvyn ja instituutioiden välistä yhteyttä tutkittaisiin tarkemmin.

Instrumenttimuuttujan valinta voi olla erityisen ongelmallista. Tästä aiheesta on tehty valtavasti tutkimuksia ja jatkotutkimusten kannalta mielenkiintoista olisi tarkastella instrumenttimuuttujien pätevyyttä. Jos instrumenttimuuttuja on vain heikosti korreloinut selittävän muuttujan kanssa, instrumenttimuuttujaestimaattorit voivat olla harhaisia (Wang & Hsiao, 2011). Tällöin instrumenttimuuttujamenetelmällä ei saada sen luotettavampia tuloksia kuin pienimmän neliösumman menetelmälläkään. Esimerkiksi koskien analysoimaani tutkimusta siirtomaiden instituutioiden vaikutuksesta taloudelliseen suorituskykyyn, voitaisiin tutkia kuolleisuuslukujen pätevyyttä instrumenttimuuttujana.

Endogeenisuusongelmaa voidaan tutkia hyvinkin laajasti. Mitä syvemmälle endogeenisuusongelman analysoinnissa edetään, sitä luotettavampia tuloksia saadaan ja sitä paremmin voidaan ymmärtää erilaisia lineaarisia yhteyksiä.

LÄHDELUETTELO

Acemoglu, D., Johnson, S. & Robinson, J. A. (2000) The Colonial Origins of Comparative Development: An Empirical Investigation. Työpaperi, National Bureau of Economic Research. Saatavilla <http://www.nber.org/papers/w7771>

Berkovitz, D. & Hoekstra, M. (2010) Does High School Quality Matter? University of Pittsburgh. Saatavilla http://www.ewi-ssl.pitt.edu/econ/files/faculty/papers/110810_pub_BerkowitzDaniel_schoolquality.pdf

Bound, J., Jaeger, D. A. & Baker, R. M. (1995) Problems with Instrumental variables Estimation When the Correlation Between the Instruments and the Endogenous Explanatory Variable is Weak. American Statistical Association 90, 430, 443-450.

Greene, W. H. (2003) Econometric Analysis. 5. p. Upper Saddle River, Prentice Hall.

Hill, R. C., Griffiths, W. E. & Lim, G. C. (2012) Principles of Econometrics. 4. p. Hoboken, John Wiley & Sons, Inc.

Roberts, M. R. & Whited, T. M. (2011) Endogeneity in Empirical Corporate Finance. Työpaperi, University of Rochester. Saatavilla http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1748604

Staiger, D. & Stock, J. H. (1997) Instrumental Variables Regression with Weak Instruments. Econometrica 65, 3, 557-586.

Wang, L. & Hsiao, C. (2011) Method of Moments Estimation and Identifiability of Semiparametric Nonlinear Errors-in-variables Models. Journal of Econometrics 165, 1, 30-44.

LIITTEET

Liite 1: Tutkimuksen aineisto ja muuttajat

Entinen siirtomaa	Bruttokansantuote 1995 (log)	Instituutiot 1995	Instituutiot 1900	Asettuminen siirtomaahan 1900	Kuolleisuus 1900
Algeria	8,39	6,50	1	0,13	78,2
Angola	7,77	5,36	1	0,08	280
Argentiina	9,13	6,39	3	0,6	68,9
Australia	9,90	9,32	7	0,98	8,55
Bahama	9,29	7,50		0,1	85
Bangladesh	6,88	5,14	1	0	71,41
Barbados	9,27		1	0,2	85
Bolivia	7,93	5,64	3	0,3	71
Brasilia	8,73	7,91	3	0,4	71
Burkina Faso	6,85	4,45	1	0	280
Chile	9,34	7,82	7	0,5	68,9
Costa Rica	8,79	7,05	7	0,2	78,1
Domini- kaaninen tasavalta	8,36	6,18	3	0,25	130
Egypti	7,95	6,77	7	0	67,8
El Salvador	6,25	5,82		0	483
Equador	8,47	6,55	3	0,3	71
Etelä-Afrikka	8,89	6,86	3	0,22	15,5
Etiopia	6,11	5,73	7	0	26
Gambia	7,27	8,27	1	0	1470
Ghana	7,37	6,27	1	0	668
Guatemala	8,29	5,14	1	0,2	71
Guinea	7,49	6,55	1	0	483
Guyana	7,90	5,89	1	0,02	32,18
Haiti	7,15	3,73	1	0	130
Honduras	7,69	5,32	5	0,2	78,1
Hong Kong	10,05	8,14		0,04	14,9
Indonesia	7,33	8,27	1	0	48,63
Intia	8,07	7,59	1	0	170
Jamaika	8,19	7,09	1	0,019	130
Kamerun	7,50	6,45	1	0	280
Kanada	9,99	9,73	7	0,99	16,1
Kenia	7,06	6,05	1	0	145
Keski-Afrikan liittovaltio	7,19		1	0	280
Kolumbia	8,81	7,32	5	0,2	71

Entinen siirtomaa	Bruttokansantuote 1995 (log)	Instituutiot 1995	Instituutiot 1900	Asettuminen siirtomaahan 1900	Kuolleisuus 1900
Kongo (Ranska)	7,42	4,68	1	0	240
Kongon demokraattinen tasavalta	6,87	3,50	1	0,08	240
Madagaskar	6,84	4,45	1	0	536,04
Malesia	8,89	7,95	1	0	17,7
Mali	6,57	4,00	1	0	2940
Malta	9,43	7,23		0	16,3
Marokko	8,04	7,09	1	0,01	78,2
Mauritania	7,41		1	0	280
Mauritius	9,05		1	0,05	30,5
Meksiko	8,94	7,50	1	0,15	71
Myanmar		5,77	1	0	34,6
Nicaragua	7,54	5,23	1	0,2	163,3
Niger	6,73	5,00	1	0	400
Nigeria	6,81	5,55	1	0	2004
Norsunluuranikko	7,44	7,00	1	0	668
Pakistan	7,35	6,05	1	0	36,99
Panama	8,84	5,91	1	0,2	163,3
Paraguay	8,21	6,95	2	0,25	78,1
Peru	8,40	5,77	3	0,3	71
Ruanda	6,48		1	0	280
Senegal	7,40	6,00	1	0	164,66
Sierra Leone	10,15	9,32	1	0,04	17,7
Singapore	7,95	5,00	3	0,2	78,1
Sri Lanka	7,73	6,05	1	0	69,8
Sudan	7,31	4,00	1	0	88,2
Suriname	8,01	4,68	1	0,01	32,18
Tansania	6,25	6,64	1	0	145
Togo	7,22	6,91	1	0	668
Trinidad ja Tobago	8,77	7,45	1	0,4	85
Tšad	6,84		1	0	280
Tunisia	8,48	6,45	1	0,03	63
Uganda	6,97	4,45	1	0	280
Uruguay	9,03	7,00	1	0,6	71
Uusi Seelanti	9,76	9,73	7	0,93	8,55
Venezuela	9,07	7,14	3	0,2	78,1
Vietnam	7,28	6,41	1	0	140
Yhdysvallat	10,22	10,00	7	0,875	15