

Mikrosimulointimallin rakentamisen, ylläpidon ja kehittämisen siirtäminen Tilastokeskukseen

Tausta

Mikrosimulointimenetelmä yksilötason aineistoihin liittyvine laskentamalleineen on tärkeä laskentaväline vero- ja sosiaaliturvalainsäädännön valmistelussa, näiden järjestelmien budjetti- ja tulonjakovaikutusten arvioinnissa sekä tavoiteltujen vaikutusten toteutumisen seurannassa. Valtiovarainministeriön ja sosiaali- ja terveysministeriön hallinnonaloilla on tähän saakka nojautunut kummallakin taholla erikseen jo 1980-luvun lopulta lähtien kehitettyihin ja jatkuvasti ylläpidettyihin TUJA- ja SOMA –mikrosimulointimalleihin. Lisäksi Kelassa on ollut muutamia vuosia kehitteillä ja käytössä JUTTA-malli. Vaikka mallit eroavat teknisesti ja osittain myös laajuudeltaan toisistaan, ovat ne menetelmällisesti yhteneväiset ja käyttävät samaa pohja-aineistoa (Tilastokeskuksen tulonjakotilaston palveluaineisto). Mallien päivitystyö ja kehittäminen on ollut lähes täysin mallien alkupeleistä tekijöiden varassa. Malleihin perehtyneitä käyttäjiä on ollut suhteellisen vähän ja näiden mallien soveltaminen valtionhallinnon ulkopuolella tutkimukseen ja selvitystyöhön on ollut vähäistä.

Lähtötilanteen taloudelliset ja demografiset muutokset vaativat mikrosimulointityökalun kehittämistä siten, että laskentamallia voidaan käyttää tehokkaammin (laajemmat aineistot) ja monipuolisemmin (koko julkisen sektorin uudelleenjakopolitiikka) apuna lainsäädännön valmistelussa sekä eri järjestelmien yhteisvaikutusten arvioinnissa. Hallinnon käyttämän mikrosimulointimallin luotettavuus ja ylläpidon jatkuvuus ovat lainsäädännön valmistelun kannalta etusijalla, mikä edellyttää sitä, että käytössä olevat mallit eivät voi olla yksittäisistä henkilöistä riippuvaisia enää tulevaisuudessa. Nämä tarpeet ja päämäärät on tiedostettu yhteisesti ja sen vuoksi on haluttu selvittää, millä edellytyksillä mikrosimulointimallin ylläpito voitaisiin siirtää Tilastokeskukselle. Samanaikaisesti on alettu rakentaa uutta, keskitettyä mallia, joka vastaisi nykyisiin ja tuleviin tarpeisiin. Tarkoituksena on, että uusi malli voisi olla myös laajemmassa käytössä ministeriöissä ja niiden alaisissa laitoksissa, yliopistoissa sekä tutkimuslaitoksissa. Hankkeessa ovat mukana valtiovarainministeriön lisäksi sosiaali- ja terveysministeriö, VATT, Kela sekä THL.

Tilastokeskuksella on parhaat edellytykset vastata mikrosimulointimallin keskitetystä tuottamisesta ja mallin jatkuvuuden turvaamisesta. Tilastokeskuksessa on asiantuntemusta, jota tarvitaan mm. aineiston ja sitä soveltavan mikrosimuloinnin tulosten analysoinnissa, tietojärjestelmien kehittämisessä ja ohjelmoinnissa. Tilastoviranomaisena Tilastokeskuksella on erityiset mahdollisuudet koota ja yhdistää henkilötason aineistoja. Tilastokeskuksella on lisäksi laajat yhteistyösuhteet valtionhallinnon eri osiin, tutkimuslaitoksiin, rekisteriaineistojen tuottajiin ym.

Keskitetyn mallin tavoitteena on rationalisoida nyt hajallaan olevaa ja työlästä, kolmen erillisen mikrosimulointimallin päivitys- ja ylläpitotyötä. Tavoitteena on toisaalta vähentää päällekkäistä työtä, toisaalta turvata aikaisempaa paremmin mallin keskeytyksetön käytettävyyttä ja toimintavarmuutta sekä riippumattomuutta yksittäisistä henkilöistä tai organisaatorakenteista. Mallien erillinen ylläpito on jarruttanut myös mikrosimulointimallien kehittämistyötä. Suomi on jäänyt jälkeen mallien kehittämisessä verrattuna muihin pohjoismaihin, joissa on kehitetty dynaamisia malleja, välittömän verotuksen ja palvelujen käytön moduuleja sekä hyödynnetty huomattavasti suurempia pohja-aineistoja.

Tietotekniikan ja aineistojen kehitys tarjoavat uusia mahdollisuuksia mikrosimuloinnin kehittämiseen. Kehittäminen vaatii kuitenkin voimavaroja ja aktiivisuutta. Mallin ylläpito ja kehittäminen vaativat jatkuvaa päivittämistä ja mallin testaamista käytännön sovelluksissa. Tähän kehitystyöhön tarvitaan useamman henkilön panosta ja tiivistä yhteistyötä käyttäjien ja pohja-aineistojen tuottajien kanssa. Lisäksi tehdään molempia tahoja hyödyttävää yhteistyötä Suomen Akatemian käynnissä olevan mikrosimulointi-mallin tutkimushankkeen kanssa.

Tavoite

Mikrosimulointi -hankkeen tavoitteena on luoda kokonaan uusi, nykyisiä kehittyneempi mikrosimulointimalli ja samalla keskittää nykyisin eri tahoilla tapahtuva päivitystyö Tilastokeskukseen. Valtionhallintoa palveleva mallien kehittäminen keskitetään Tilastokeskukseen. Uuden mallin ensimmäisen käyttöön soveltu-
van version on tarkoitus valmistua vuoden 2011 loppuun mennessä.

Uuden mallin luominen alusta pitäen mahdollistaa nykyisiä malleja kehittyneemmän rakenteen, johon voidaan myöhemmin sovittavalla tavalla sisällyttää uusia mallinnettavia osa-alueita, kuten välillinen verotus, ansioeläkejärjestelmä ja hyvinvointipalvelut. Myös dynaamisia osioita, kuten työntarjonta ja pidemmän aikavälin vaikutukset voidaan liittää malliin. Mallista tehdään helposti omaksuttava, jolloin sen käyttäjäkunta voi kasvaa ja sovellusalueet laajentua. Lisäksi laskentatarkkuutta parannetaan suuremman väestöä edustavan pohja-aineiston avulla.

Hankkeen eteneminen

Työryhmä ja seurantaryhmä

Valtiovarainministeriö käynnisti vuoden 2010 alusta hankkeen uuden mikrosimulointimallin ylläpidon ja kehittämisen siirtämisestä Tilastokeskukseen. Hankkeen etenemistä varten perustettiin työryhmä, jonka jäseninä ovat valtiovarainministeriöstä Elina Pylkkänen (pj.), Pertti Honkanen (siht.), Tuulia Hakola-Uusitalo (tilalla: Laura Vartia), Fransiska Pukander, Merja Taipalus ja Marja Tuovinen; sosiaali- ja terveysministeriöstä Ilari Keso, Kelasta Anita Haataja, Terveystieteiden ja hyvinvoinnin laitoksesta Seppo Sallila; Valtion taloudellisesta tutkimuskeskuksesta Heikki Viitamäki sekä Tilastokeskuksesta Juha Honkkila (31.12.2010 asti), Antti Katainen (1.1.2011 lähtien) ja Aila Äyväri. Työryhmän tehtävänä on olla mukana suunnittelemassa ja seuraamassa toteutettavan mikrosimulointimallin perusrakennetta. Työryhmä tutustuu tarkemmin erityisesti muiden Pohjoismaiden mallien organisointitapoihin ja itse malleihin esim. kutsumalla asiantuntijoita työryhmän kuultavaksi. Työryhmä kokoontuu säännöllisesti (tähän mennessä ollut 20 kokousta) ja raportoi säännöllisin väliajoin edistymisestään hankkeen seurantaryhmälle, joka on kokoontunut tähän mennessä kahdesti. Seurantaryhmään kuuluvat puheenjohtajana Lasse Arvela (VM) ja jäseninä Tuomas Sukselainen (VM), Kari Välimäki (STM) lopettanut 6.4.2011: tilalle Raija Volk (STM), Seija Ilmakunnas (VATT) lopettanut 1.11.2010, tilalle Aki Kangasharju (VATT), Olli Kangas (Kela), Pasi Moisio (THL), Riitta Harala (TK) ja Jarmo Hyrkkö (TK).

Uuden mallin rakentaminen

Uuden mikrosimulointimallin toteuttajaksi ja kehittäjäksi valtiovarainministeriö on palkannut Pertti Honkasen ajalle 1.2.2010 – 31.12.2011. Honkanen työskentelee mallin rakenteen suunnittelu- ja toteutusvaiheen aikana Tilastokeskuksessa, jossa pystyy parhaiten testaamaan uusia aineistoratkaisuja ja tuotantoympäristöä. Mallin toteutuksessa otetaan huomioon muissa maissa tehty ratkaisut ja kokemukset (Honkasen raportti aiheesta). Mallin rakentamisessa pyritään soveltamaan uusimpia teknisiä ratkaisuja. Mukaan uuden mikrosimulointimallin kehitykseen voidaan tarvittaessa pyytää ohjelmistojen tuottajien edustajia.

Tilastokeskuksen esiselvitysprojekti

Tilastokeskuksessa toteutettiin mikrosimulointimallin ylläpitovastuun siirtämisen edellytyksiä pohtinut esiselvitysprojekti 1.2. – 30.4.2010. Esiselvitysprojektin tarkoituksena oli selvittää mallin vaatimia tietotekniikkaratkaisuja ja kartoittaa mallin keskeiset käyttäjä- ja osaamistarpeet sekä määrittää Tilastokeskukselle siirrettävät vastuut. Selvityksen tehtävänä oli myös kartoittaa tehtävän hoitamisesta aiheutuva rahoituksen ja henkilöstön tarve. Projektin loppuraportti julkaistiin 15.5.2010 (Juha Honkkila) ja esiselvityksen tulokset esiteltiin hankkeen seurantaryhmälle 12.5.2010 pidetyssä kokouksessa.

Työryhmässä vierailleet asiantuntijat

Työryhmä on kuullut seuraavia asiantuntijoita mikrosimulointi-hankkeeseen liittyen:

TUJA-, SOMA-, JUTTA- ja HVS-mallien esittelyt

prof. Jukka Pirttilä, Suomen akatemian rahoittaman mikrosimulointi-hankkeen johtaja (18.2.2010);

Ruotsin tilastokeskuksen FASIT-mikrosimulointimallin tuotannosta vastaavat Leif Johansson ja Klas Lindström (7. – 8.4.2010 valtiovarainministeriössä ja Tilastokeskuksessa);
Norjan tilastokeskuksen Mozart-mallin asiantuntija Nils Martin Stølen (26.11.2009);
Tilastokeskuksen tutkimuspalveluyksikön toimintaa esitelleet Jussi Heino ja Antti Katainen (8.9.2010);
SAS Institute Oy esitteli tarjoamaansa ohjelmistoa ja käyttöliittymävaihtoehtoja (12.10.2010)
Suomen Pankin kokonaistasapainomallin, Aino-mallin, esittelijänä Jukka Railavo (29.10.2010)
Eläketurvakeskuksen tutkimusosaston päällikkö Mikko Kautto, esitteli mm. eläkekertymätietojen aineistopohjia (18.4.2011)

Mikrosimulointi –hankkeen taloudellinen ja työvoimaresursointi

Hankkeen työryhmä ja seurantaryhmä työskentelevät virkatyönä ilman eri korvausta.

Valtiovarainministeriö on palkannut määrä- ja osa-aikaisesti 1.2.2010 – 31.12.2011 (80 %) Pertti Honkasen rakentamaan uutta mikrosimulointimallia.

Tilastokeskuksen tekemän esiselvitysprojektin laatimisesta valtiovarainministeriö suoritti korvauksen 15 000 euroa Tilastokeskukselle.

Ruotsin tilastokeskuksen mikrosimulointi-asiantuntijoiden vierailun kustannuksista vastasi valtiovarainministeriö n. 1 000 euroa.

Täydentävässä talousarviossa (lokakuu 2010) myönnettiin mikrosimulointihankkeelle lisärahoitusta, jonka mukaan Tilastokeskus saa 150 000 euroa vuodelle 2011. Tämä lisäresurssi annettiin mikrosimulointihankkeen uusia rekrytointejä varten. Rahasummalla pystytään palkkaamaan kolme henkilöä mikrosimulointimallin tuotantotehtävään.

Kehysbudjetissa on esitetty Tilastokeskukselle vuosille 2012 – 2015 toimintamenoihin 200 000 euroa vuosittain mikrosimuloinnin kehittämiseen.

Hankkeessa mukana olevilta tahoilta (Kela, THL ja VATT) on saatu erityisesti lupauksia asiantuntija-avun antamisesta hankkeen edetessä. Kela on perustanut mikrosimulointiin keskittyvän tutkimusyksikön, jota johtaa Pertti Honkanen.

Tilastokeskus tulee suunnitelmien mukaan myymään mikrosimulointi –mallia eri käyttäjätahoille ja saa siten myyntituloja mallin myynnistä, konsultaatiosta sekä selvitysten tekemisestä.

Mallin substanssi

Simulointimallin on sovelluttava sekä esimerkkilaskelmiin että varsinaiseen simulointiin. Ensi vaiheessa malli kattaisi lainsäädännön vähintään nykyisten mallien laajuudessa. Perusosiot ovat silloin:

- tuloverotus;
- sairausvakuutuksen päivärahat;
- työttömyysturvan päivärahat;
- kansaneläkejärjestelmään kuuluvat etuudet;
- lasten kotihoidon tuki ja päivähoidomaksut;
- lapsilisät ja elatustuet;
- opintotuki ja aikuisopintotuki;
- yleinen ja eläkkeensaajien asumistuki;
- toimeentulotuki.

Täydentäviä osioita voivat esimerkiksi olla työnantajamaksut sekä sosiaaliturvan rahoitusta koskevat säännökset sekä työeläkelainsäädäntö esimerkkilaskelmia varten. Lakimallia voidaan täydentää työvoiman tarjontaa koskevalla osiolla. Erikseen tutkitaan kulutusverotuksen mallintamista, mikä edellyttää kulutusaineiston lisäämistä malliaineistoon. Tässä vaiheessa ei ole tarkoitus ottaa kantaa kehittämistehtävien priorisointiin.

Lainsäädäntö koodataan pidemmältä ajalta esim. pääosin 1990-luvun alusta lähtien. (Pidemmällä aikavälillä etuus- ja lakimallit liitetään myös vanhempiin aineistoihin mm. seurantatutkimusta varten.)

Simulointimallissa automatisoidaan keskeisten yhteenvedo- ja vertailutaulukoiden luonti. Kun lasketaan käytettävissä olevia tuloja koko väestölle, tuotetaan myös keskeiset tulonjakoindeksit, ainakin Gini-kerroin ja köyhyysasteet. Käyttäjillä säilyy kuitenkin yhteys yksilötason lähtö- ja tulostiedostoihin jatko-analyysia varten. Esimerkkilaskelmissa on mahdollistettava erilaisten taulukoiden luominen kuvaamaan etuuksien ja verotuksen riippuvuutta eri muuttujista ja toisistaan.

Simulointimalliin liitetään ajantasaistaminen ja uudelleenpainotus siten, että simulointi uusinta aineistovuotta uudempien vuosien aineistolla sekä ennusteita hyväksikäyttäen olisi mahdollisimman realistista budjettilaskelmia tehtäessä. Tavoitteena on, että eri käyttäjillä on valmistelutyössä yhteinen malli- ja aineistopohja. Tällä tavoin voidaan vähentää sekaannuksia ja turhaa selvittelytyötä, jota ei aineistopohjilla tehdyt arviot saattavat aiheuttaa.

Tekniset vaatimukset

Teknisesti mallilta edellytetään tehokkuutta, helppokäyttöisyyttä ja koodin avoimuutta. Se saavutetaan käyttämällä ohjelmointivälineitä pääosin:

- SAS-tilasto-ohjelmaa;
- Excel-tilulukkolaskentaa
- C-kieltä

Tässä vaiheessa käytettäviä ohjelmointivälineitä koskevat täsmälliset ratkaisut ovat auki. Ohjelmakoodin laatuun on kiinnitettävä huomiota, koska mallilla tulee olemaan useita käyttäjiä, joiden tulee ymmärtää koodin sisältö ja pystyä sitä muuntamaan. Koodin ymmärtävyyden parantamiseksi muuttujien, funktioiden, parametrien, tiedostojen ym. nimiä standardisoidaan.

Käyttäjä voi vaikuttaa laskelmiin simuloinnin oletuksia ja lainsäädännön parametreja muuttamalla sekä lainsäädännön laskukaavoja muokkaamalla. Käyttäjä voi laatia valmiiksi ohjelmoitujen yhteenvedotaulukoiden lisäksi omia erilaisiin luokitteluihin perustuvia taulukoita ja erittelyjä (=yhteys yksilötasoisii tulostiedostoihin).

Lainsäädäntö ohjelmoidaan luomalla johdonmukaiset parametritaulukot sekä ohjelmoimalla laskukaavat systemaattisesti funktioiksi tai makroiksi. Tässä voidaan hyödyntää SAS-makroja, SAS 9.2-version funktio-toimintoa, C-kielillä koodattuja funktioita sekä täydentävästi lähinnä esimerkkilaskelmissa myös Visual Basic -funktioita. Funktio- ja makromuoto parantaa laskukaavojen siirrettävyyttä eri aineistoihin, eri tehtäviin ja eri ohjelmiin. Toisaalta sinällään tehokkaat ohjelmointiratkaisut voivat vaativuudellaan nostaa mallin käyttökynnystä ja rajoittaa sen käyttöä.

Nykyisenkaltaiseen aineistoon pohjautuva malli asennetaan aineistoihin eri käyttäjien tai organisaatioiden henkilökohtaisille koneille tai palvelimille. Suuremman rekisteripohjaisen aineiston etäkäyttöä tutkitaan. Mallien ajamiseen tarvittavat tiedostot jaetaan Tilastokeskuksen palvelimen kautta eri käyttäjille, mutta kuitenkin käyttäjä vastaa itse tarvittavien ohjelmien (SAS, Excel ym.) hankkimisesta.

Aineisto

Uusi malli rakennetaan aluksi nykyisenkaltaisen tulonjakoa-aineiston (ns. palveluaineisto) pohjalle, mutta samalla aletaan selvittää laajemman datan tuottamista ja käyttöä mallintamisessa. Laajemmasta datasta käytetään hyväksi samoja rekisteriaineistoja, joita on käytetty ns. palveluaineiston tuottamiseen. Selvitetään mahdollisuuksia liittää aineistoon nykyistä enemmän mm. työvoimahallinnon, Eläketurvakeskuksen, Kelan ja THL:n rekisteritietoja. Sellaisia tietoja, jotka perustuvat haastatteluun, imputoidaan mahdollisuuksien mukaan laajempaan aineistoon. Selvittettäviä kysymyksiä ovat mm. uuden otoksen koko, paneeliominaisuus, laitosväestön osuus ym. Myös mahdollisuuksia kokonaisaineiston hyväksikäyttöön joissakin simuloinneissa selvitetään.

Organisaatio ja eri organisaatioiden vastuu

Mallin ylläpito sijoitetaan Tilastokeskuksen tutkimuspalveluyksikköön. Tilastokeskus vastaa aineiston tuottamisen ohella aina ajantasaisen ja teknisesti toimivan mallin tuottamisesta ja saatavilla olemisesta. Tilastokeskuksen vastuulle kuuluu myös käyttäjätuki. Lainsäädännön ohjelmoinnissa käytetään kuitenkin hyväksi mallia käyttävien organisaatioiden asiantuntemusta (esim. verolainsäädännössä vm:n, sosiaaliturvalainsäädännössä STM:n, Kelan, THL:n jne.) Myös aineiston tuottamiseen liittyvistä kysymyksistä kuten eri muuttujien sisällöstä ja tarpeellisuudesta Tilastokeskus konsultoi mallin käyttäjäorganisaatioita.

Tilastokeskus vastaa myös ajantasaistuksen ja uudelleenpainotuksen toimivuudesta. Ajantasaistetun aineiston tuottamisessa käytetään pohjana vm:n ennusteita taloudellisesta kehityksestä. Eri organisaatiot käyttävät mallia omalla vastuullaan. Mallien käyttövarmuuden ja työn jatkuvuuden kannalta on tarpeen, että Tilastokeskuksen ohella ainakin tärkeimmissä mallia käyttävissä organisaatioissa on useampia mallin käyttöön ja kehittämiseen perehtyneitä henkilöitä.

Dokumentaatio

Malli dokumentoidaan huolellisesti, mikä palvelee mallin käyttäjiä ja kehittäjiä. Dokumentaatioon kuuluvat:

- käyttöohjeet/käsikirja;
- lainsäädännön parametrien ja laskukaavojen (funktioiden, makrojen) ja tulokäsitteiden dokumentointi;
- simulointiohjelman eri vaiheiden dokumentointi;
- ohjelmakoodin dokumentointi koodin ohkeen kirjoitetuilla kommentteilla;
- aineiston dokumentointi siten kuin Tilastokeskus jo dokumentoi tulonjakotilaston palveluaineistoa.
-

Koulutus

Tilastokeskus järjestää mallin käyttäjille koulutusta, jossa perehdytään malliaineistoon, mallin rakenteeseen ja käyttöön sekä myös mallissa tarvittaviin ohjelmointivälineisiin.

Käyttäjäkustannukset

Tilastokeskus velottaa luovutetusta mallista maksun, jonka suuruutta ei ole vielä määritetty. Tämän lisäksi käyttäjät tarvitsevat SAS-ohjelmistolisenssin (SAS Analytics Pro + SAS Access to PC File Formats), jonka hinta on ensimmäisenä vuonna 2 400 € + ALV ja jatkovuosimaksu 931 € + ALV. Mallin pohja-aineistona käytettävän Tulonjakotilaston palveluaineiston kustannus on tällä hetkellä 6 410 € / vuosi.

Kansainvälinen yhteistyö

Tilastokeskus ja muut mallityöhön osallistuvat organisaatiot seuraavat alan kehitystä mm. osallistumalla Pohjoismaisiin mikrosimulointikokouksiin, kansainvälisen mikrosimulointijärjestön (IMA) työhön sekä EUROMOD-hankkeen kehittämiseen (Heikki Viitamäki, VATT).

Vuoden 2010 pohjoismaisessa kokouksessa Köpenhaminassa mikrosimulointimalleja hyödyntäviä suomalaisia tutkimuksia esittelivät Anita Haataja (Kela) sekä Pertti Honkanen (Kela, VM).

Vuoden 2011 kansainvälinen mikrosimulointi-konferenssi järjestetään Tukholmassa 8. – 10.6.2011. Suomesta konferenssiin osallistuu työryhmän jäseniä sekä ensimmäistä kertaa myös Tilastokeskuksen mikrosimulointi-projektiin kuuluvat henkilöt.